

AJS

American

Journal

Sociology

Volume 105 Number 4

January 2000

Collective Memory in Russia——Schuman and Corning

*Working-Class Power, Capitalist-Class Interests, and Class
Compromise*——Wright

*Strategic Capacity in the Unionization of California Agriculture,
1959–1966*——Ganz

The Outcomes of Homeless Mobilization——Cress and Snow

The Emergence of Generalized Exchange——Takahashi

*Economic Transformation and Income Inequality in Urban
China*——Zhou

Controversies and Evidence——Cao and Nee

The University of Chicago Press

KP-~~893~~ 893
ROGER V. GOULD, Editor

EMILY BARMAN, JIM DAVIS, ANDREAS GLAESER, TONY PAIK, JENNIFER PASHUP, and JESPER SØRENSEN, Associate Editors

LESLIE SALZINGER, Associate Editor, Book Review

KELLY DALEY, DAVID GRAZIAN, JULIA GWYNNE, SUKARI IVESTER, HARRIS KIM, CHI-YOUNG KOH, KWAI HANG NG, and STEPHEN ROSENBERG, Book Review Editorial Board

SUSAN ALLAN, Managing Editor

CONNIE FRITSCHKE and GRETCHEN KALWINSKI, Editorial Assistants

Consulting Editors VICTORIA D. ALEXANDER • CHRISTOPHER ANSELL • NICOLA BEISEL • YANJIE BIAN • WILLIAM BRUSTEIN • STEVEN CLAYMAN • SAMUEL COHN • FRANK DOBBIN • SCOTT ELIASON • MUSTAFA EMIRBAYER • WENDY NELSON ESPELAND • MYRA MARX FERREE • JACK GOLDSTONE • PHILIP GORSKI • GUANG GUO • LINGXIN HAO • KAREN HEGTVEDT • BARBARA KILBOURNE • RUUD KOOPMANS • JOHN LAUB • EDWARD J. LAWLER • MICHAEL LOVAGLIA • SAMUEL R. LUCAS • MICHAEL MACY • ROSS MATSUEDA • DEBRA C. MINKOFF • CAROL MUELLER • JOANE P. NAGEL • HAYAGREEVA RAO • WILLIAM G. ROY • DIETRICH RUESCHEMEYER • JANET CHAFETZ SALTZMAN • PAMELA SMOCK • SCOTT SOUTH • MARC STEINBERG • TONY TAM • BRIAN UZZI • BRUCE WESTERN • MICHAEL WHITE • RAYMOND WONG

Editorial Board ANDREW D. ABBOTT • GARY S. BECKER • CHARLES E. BIDWELL • DONALD J. BOGUE • ANTHONY BRYK • RONALD BURT • TERRY N. CLARK • JOHN L. COMAROFF • LARRY HEDGES • PATRICK HEUVELINE • EVELYN KITAGAWA • EDWARD O. LAUMANN • DONALD N. LEVINE • WILLIAM L. PARISH • MARTIN RIESEBRODT • ROBERT J. SAMPSON • SASKIA SASSEN • ROSS M. STOLZENBERG • FRED L. STRODTBECK • GERALD D. SUTTLER • RICHARD TAUB • LINDA J. WAITE • KIM WEEDEN • KAZUO YAMAGUCHI • DINGXIN ZHAO

AMERICAN JOURNAL OF SOCIOLOGY (ISSN 0002-9602) is published bimonthly (every other month) in July, September, November, January, March, and May by the University of Chicago Press, Journals Division, 5720 S. Woodlawn Ave., Chicago, Illinois 60637. Subscription rates per year: U.S. institutions, \$153.00; individuals, \$43.00; students, \$30.00 (photocopy of valid student ID must accompany subscription); ASA and BSA members, \$38.00. All other countries add \$12.00 to cover postage. Canadian subscribers add 7% GST. Subscription agent for Japan: Kinokuniya Company, Ltd. Single copy rates: institutions \$25.50, individuals \$7.15. Volumes 1-94 available from Periodical Services Co., 11 Main St., Germantown, New York 12526. All volumes are available from University Microfilms International, 300 North Zeeb Road, Ann Arbor, Michigan 48106. Subscriptions are payable in advance and will begin with the first issue published after receipt of order. Checks should be made payable to *American Journal of Sociology*. For subscription information, write University of Chicago Press, Journals Division, P.O. Box 37005, Chicago, Illinois 60637; fax, 312-753-0811; or E-mail subscriptions@journals.uchicago.edu (editorial correspondence, E-mail AJS@cicero.uchicago.edu).

Postmaster: Send address changes to *American Journal of Sociology*, University of Chicago Press, Journals Division, P.O. Box 37005, Chicago, Illinois 60637.

Copying beyond Fair Use. The code on the first page of an article in this journal indicates the copyright owner's consent that copies of the article may be made beyond those permitted by Section 107 or 108 of the U.S. Copyright Law provided that copies are made only for personal or internal use or for the personal or internal use of specific clients and provided that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, Massachusetts 01923. To request permission for other kinds of copying, such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale, kindly write to Permissions Department, University of Chicago Press, 5801 South Ellis Avenue, Chicago, Illinois 60637. If no code appears on the first page of an article, permission to reprint may be obtained only from the author.

Periodicals postage paid at Chicago, Illinois, and at additional mailing offices.

© 2000 by The University of Chicago. All rights reserved.

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984. (∞)

KP-~~8~~ 893

COR-H05423-16-12893

AJS

American Journal of Sociology

Volume 105 No. 4 January 2000

CONTENTS

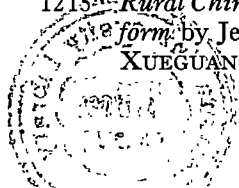
- 913 Collective Knowledge of Public Events: The Soviet Era from the Great Purge to Glasnost
HOWARD SCHUMAN AND AMY D. CORNING
- 957 Working-Class Power, Capitalist-Class Interests, and Class Compromise
ERIK OLIN WRIGHT
- 1003 Resources and Resourcefulness: Strategic Capacity in the Unionization of California Agriculture, 1959-1966
MARSHALL GANZ
- 1063 The Outcomes of Homeless Mobilization: The Influence of Organization, Disruption, Political Mediation, and Framing
DANIEL M. CRESS AND DAVID A. SNOW
- 1105 The Emergence of Generalized Exchange
NOBUYUKI TAKAHASHI
- 1135 Economic Transformation and Income Inequality in Urban China: Evidence from Panel Data
XUEGUANG ZHOU
- 1175 Comment: Controversies and Evidence in the Market Transition Debate
YANG CAO AND VICTOR G. NEE
- 1189 Reply: Beyond the Debate and toward Substantive Institutional Analysis
XUEGUANG ZHOU

16



Book Reviews

- 1196 *ReOrient: Global Economy in the Asian Age* by Andre Gunder Frank
CHRISTOPHER CHASE-DUNN
- 1198 *Nationalism, Liberalism, and Progress. Vol. 1, The Rise and Decline of Nationalism* by Ernst B. Haas
JOHN A. ARMSTRONG
- 1200 *Fantasy City: Pleasure and Profit in the Postmodern Metropolis* by John Hannigan
DAVID GRAZIAN
- 1202 *Globalisation, Human Rights, and Labour Law in Pacific Asia* by Anthony Woodiwiss
PHENG CHEAH
- 1204 *Los Angeles: Globalization, Urbanization and Social Struggles* by Roger Keil
JAN LIN
- 1206 *The Rights Revolution: Lawyers, Activists, and Supreme Courts in Comparative Perspective* by Charles R. Epp
SIDNEY TARROW
- 1207 *Rescuing Business: The Making of Corporate Bankruptcy Law in England and the United States* by Bruce G. Carruthers and Terence C. Halliday
JOHN HENRY SCHLEGEL
- 1209 *Worlds of Production: The Action Frameworks of the Economy* by Michael Storper and Robert Salais
GERALD BERK
- 1211 *Tracing the Veins: Of Copper, Culture, and Community from Butte to Chuquicamata* by Janet L. Finn
BARBARA SUTTON
- 1213 *Manufacturing Ideology: Scientific Management in Twentieth-Century Japan* by William M. Tsutsui
GARY HERRIGEL
- 1215 *Rural China Takes Off: Institutional Foundations of Economic Reform* by Jean C. Oi
XUEGUANG ZHOU



- 1217 *China's Catholics: Tragedy and Hope in an Emerging Civil Society*
by Richard Madsen
DANIEL H. BAYS
- 1219 *Forever Foreigners or Honorary Whites? The Asian Ethnic Experience Today* by Mia Tuan
GRACE KAO
- 1221 *Intimate Ironies: Modernity and the Making of Middle-Class Lives in Brazil* by Brian P. Owensby
MIGUEL ANGEL CENTENO
- 1222 *Worlds Apart: Why Poverty Persists in Rural America* by Cynthia M. Duncan
DAVID BROWN
- 1224 *The Silent War: Imperialism and the Changing Perception of Race*
by Frank Füredi
JOHN WILLINSKY
- 1226 *Why Americans Hate Welfare: Race, Media, and the Politics of Antipoverty Policy* by Martin Gilens
SANFORD F. SCHRAM
- 1228 *Postmodernity and the Fragmentation of Welfare*, edited by John Carter
C. G. A. BRYANT
- 1230 *States, Markets, Families: Gender, Liberalism and Social Policy in Australia, Canada, Great Britain and the United States* by Julia S. O'Conner, Ann Shola Orloff, and Sheila Shaver
LISA D. BRUSH
- 1232 *What Difference Does a Husband Make? Women and Marital Status in Nazi and Postwar Germany* by Elizabeth D. Heineman
LYNNE HANEY
- 1234 *At the Breast: Ideologies of Breastfeeding and Motherhood in the Contemporary United States* by Linda M. Blum
BARBARA KATZ ROTHMAN
- 1236 *The Body in Everyday Life*, edited by Sarah Nettleton and Jonathan Watson
BARRY GLASSNER
- 1238 *Mema's House, Mexico City: On Transvestites, Queens, and Machos* by Annick Prieur
MIGUEL DÍAZ-BARRIGA

- 1239 *Gay Politics, Urban Politics: Identity and Economics in the Urban Setting* by Robert W. Bailey

Culture Wars and Local Politics, edited by Elaine B. Sharp
RHYS H. WILLIAMS

- 1242 *The Boundaries of Blackness: AIDS and the Breakdown of Black Politics* by Cathy J. Cohen
SHARON HICKS-BARTLETT

- 1244 *Fast Lives: Women Who Use Crack Cocaine* by Claire E. Sterk
KATHRYN J. FOX

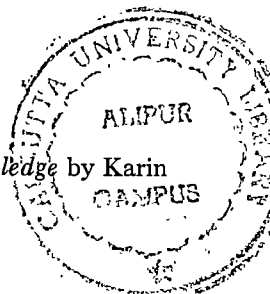
- 1246 *Inventing Pain Medicine: From the Laboratory to the Clinic* by
Isabelle Baszanger
NICHOLAS A. CHRISTAKIS

CONTENTS

- 1249 The Instability of Androgynous Names: The Symbolic Maintenance of Gender Boundaries
STANLEY LIEBERSON, SUSAN DUMAIS, AND SHYON BAUMANN
- 1288 Social Capital at Work: Networks and Employment at a Phone Center
ROBERTO M. FERNANDEZ, EMILIO J. CASTILLA, AND PAUL MOORE
- 1357 The Power to Punish: Discretion and Sentencing Reform in the War on Drugs
RODNEY L. ENGEN AND SARA STEEN
- 1396 Risk and Trust in Social Exchange: An Experimental Test of a Classical Proposition
LINDA D. MOLM, NOBUYUKI TAKAHASHI, AND GRETCHEN PETERSON
- 1428 The Mosaic Moment: An Early Modernist Critique of Modernist Theories of Nationalism
PHILIP S. GORSKI

Book Reviews

- 1469 *How to Have Theory in an Epidemic: Cultural Chronicles of AIDS*
by Paula A. Treichler
LISA JEAN MOORE
- 1471 *Lessons from the Intersexed* by Suzanne J. Kessler
BERNICE L. HAUSMAN
- 1472 *Epistemic Cultures: How the Sciences Make Knowledge* by Karin Knorr-Cetina
TREVOR PINCH



- 1475 *A Structural Theory of Social Influence* by Noah E. Friedkin
STANLEY WASSERMAN
- 1477 *On Line and On Paper: Visual Representations, Visual Culture, and Computer Graphics in Design Engineering* by Kathryn Henderson
JOHN LAW
- 1478 *Impersonal Influence: How Perceptions of Mass Collectives Affect Political Attitudes* by Diana C. Mutz
LYNN M. SANDERS
- 1480 *Identity without Selfhood: Simone de Beauvoir and Bisexuality* by Mariam Fraser
JANET WIRTH-CAUCHON
- 1482 *Fighting to Become Americans: Jews, Gender, and the Anxiety of Assimilation* by Riv-Ellen Prell
DEBORAH A. GERSON
- 1484 *Handbook of the Sociology of Gender* by Janet Saltzman Chafetz
MYRA MARX FERREE
- 1486 *Equality by Design: The Grand Experiment in Destratification in Socialist Hungary* by Szonja Szelényi
PETR MATEJU
- 1488 *Where the World Ended: Re-unification and Identity in the German Borderland* by Daphne Berdahl
ANDREAS GLAESER
- 1490 *Subnational Politics and Democratization in Mexico*, edited by Wayne A. Cornelius, Todd A. Eisenstadt, and Jane Hindley
JONATHAN FOX
- 1492 *Women and Democracy: Latin America and Central and Eastern Europe*, edited by Jane S. Jaquette and Sharon L. Wolchik
CAROL MUELLER
- 1494 *The Empowerment of Women in India: Grassroots Women's Networks and the State* by Sangeetha Purushothaman
NANCY A. NAPLES AND MANISHA DESAI
- 1496 *Fields of Protest: Women's Movements in India* by Raka Ray
VERTA TAYLOR

- 1498 *Gender and the South China Miracle: Two Worlds of Factory Women* by Ching Kwan Lee
DIANE L. WOLF
- 1500 *Contesting Citizenship in Urban China: Peasant Migrants, the State, and the Logic of the Market* by Dorothy J. Solinger
ZAI LIANG
- 1502 *From Education to Work: Cross-National Perspectives*, edited by Walter R. Heinz
KATHRYN BORMAN
- 1504 *State-Making and Labor-Movements: France and the United States, 1876-1914* by Gerald Friedman
KENNETH H. TUCKER, JR.
- 1506 *Trade Conditions and Labor Rights: U.S. Initiatives, Dominican and Central American Responses* by Henry J. Frundt
ANDREW SCHRANK
- 1508 *Forced Choices: Class, Community, and Worker Ownership* by Charles S. Varano
VICKI SMITH
- 1510 *Steelworker Alley: How Class Works in Youngstown* by Robert Bruno
CHARLES S. VARANO
- 1512 *Working Hard and Making Do: Surviving in Small Town America* by Margaret K. Nelson and Joan Smith
ANN R. TICKAMYER
- 1514 *Managing to Make It: Urban Families and Adolescent Success* by Frank F. Furstenberg, Jr., Thomas D. Cook, Jacquelynne Eccles, Glen H. Elder, Jr., and Arnold Sameroff
JOYCE LADNER
- 1516 *Chinese Christians in America: Conversion, Assimilation, and Adhesive Identities* by Fenggang Yang
REBECCA KIM AND MIN ZHOU
- 1518 *Culture on the Margins: The Black Spiritual and the Rise of American Cultural Interpretation* by Jon Cruz
MICHAEL ROGIN
- 1519 *Refried Elvis: The Rise of the Mexican Counterculture* by Eric Zolov
PABLO VILA

IN THIS ISSUE

STANLEY LIEBERSON recently completed an extensive study of the internal mechanisms driving changes in fashion and culture (with research on names as his primary data source). *A Matter of Taste: How Names, Fashions, and Culture Change* will be published late this summer by Yale University Press. He is the Abbott Lawrence Lowell Professor at Harvard University.

SUSAN DUMAIS is a doctoral candidate in sociology at Harvard University. Her dissertation employs a person-centered research strategy to analyze survey data on the determinants of working-class students' educational success. She is also interested in evaluating the effects of gender differences in cultural capital.

SHYON BAUMANN is a doctoral candidate in sociology at Harvard University. His dissertation examines the nature of cultural authority through an analysis of the influence of film critics. He is also interested in examining the ideological and organizational foundations of the valuation of art.

ROBERTO M. FERNANDEZ is a sociologist and professor of organizational behavior at the Graduate School of Business, Stanford University. He works in the areas of organizational sociology, social networks, race and ethnic relations, and social stratification. He has done extensive field research in organizations, including an exhaustive five-year case study of a plant retooling and relocation. He is continuing his research on networks and hiring by studying financial services, market research, manufacturing, and retail sales jobs.

EMILIO J. CASTILLA is a Ph.D. candidate in sociology at Stanford University. His main research interests are formal organizations, economic sociology, and comparative sociology. Currently, he studies the influence of social networks on employee performance within one organization and is involved in a network analysis of venture capital firms in Silicon Valley. His most recent book is on longitudinal methods in sociological research.

PAUL MOORE is a Ph.D. candidate in organizational behavior at Stanford University, Graduate School of Business. His dissertation is on the consequences of workplace mentoring and the mechanisms driving these effects. In addition to studying the individual and organizational value of social networks, he studies learning and attribution after accidents and other notable events.

RODNEY L. ENGEN is assistant professor of sociology in the Department of Sociology and Anthropology at North Carolina State University. His current research interests include the role of social and organizational contexts in sentencing, racial disparities, and the implementation of sentencing reforms.

SARA STEEN is assistant professor of sociology at Vanderbilt University. She is currently working, with George Bridges, on the role of typification of minority youth as a mediating factor in the relationship between race

and juvenile justice outcomes. She is also continuing work, with Rodney Engen and Randy Gainey, on the question of whether and how sentencing guidelines affect the use of discretion in cases involving felony drug offenders.

LINDA D. MOLM is professor of sociology and department head at the University of Arizona. Her primary research interest is the experimental analysis of theories of social exchange and power. She is currently coeditor of *Social Psychology Quarterly*.

NOBUYUKI TAKAHASHI is a Ph.D. candidate in the sociology department at the University of Arizona. His areas of interest include social psychology, group processes, and social networks. He is currently interested in the relationship between social networks and the type of comparison that people use in making fairness judgments.

GRETCHEN PETERSON is a Ph.D. candidate in sociology at the University of Arizona. Her areas of interest include social psychology and social stratification. She is currently researching differences in perceptions of justice between positively and negatively connected exchange networks.

PHILIP S. GORSKI is assistant professor of sociology at the University of Wisconsin—Madison. His research focuses on the interplay between religion and politics in early modern and modern Europe. He is currently working on three interrelated projects: a book entitled *The Disciplinary Revolution: Calvinism, Confessionalism, and State Formation in Early Modern Europe* (University of Chicago Press, in press); a series of essays reassessing the Weber thesis and its various offshoots; and several articles on secularization.

Information for Contributors

Editorial Procedures: All papers deemed appropriate for *AJS* are sent out anonymously to readers. The editors rely heavily on the judgments of those readers, although they are not bound by them. To protect anonymity, only the title should appear on the manuscript. Attach a cover page with the title of the manuscript, your name, and your affiliation. Manuscripts are accepted for publication subject to the understanding that *AJS* has the right of first publication. Submission of a manuscript to *AJS* is taken to indicate the author's commitment to publish in this *Journal*. No paper known to be under consideration by any other journal will be reviewed by *AJS*. In addition, *AJS* is not a suitable vehicle for materials published elsewhere, in whole or in part. Upon publication, all rights, including subsidiary rights, are owned by the University of Chicago Press. The author retains the right to use his or her article without charge in any book of which he or she is the author or editor after it has appeared in the *Journal*.

Manuscript Acceptance Policy: While it is our policy to require the assignment of copyright on most journal articles and review essays, we do not usually request assignment of copyright for other contributions. Although the copyright to such a contribution may remain with the author, it is understood that, in return for publication, the *Journal* has the nonexclusive right to publish the contribution and the continuing right, without limit, to include the contribution as part of any reprinting of the issue and/or volume of the *Journal* in which the contribution first appeared by any means and in any format, including computer-assisted storage and readout, in which the issue and/or volume may be reproduced by the publisher or by its licensed agencies.

Preparation of Copy

1. Type *all copy*—including indented matter, footnotes, tables, and references—*double spaced*, allowing generous margins at the top, bottom, and sides of page. Please do not break words at ends of lines or justify right-hand margins. Do not use proportional spacing fonts. Indicate italics by underlining only. Number all footnotes, beginning with the acknowledgment note as number 1. *Provide a word count on the cover page of the manuscript.*

2. Begin each table on a separate page. Refer to each table in numerical order in the text. Tables should not contain vertical lines or more than 20 two-digit columns or the equivalent.

3. Clarify all mathematical symbols (e.g., Greek letters) with words in the margins of the manuscript.

4. Original or glossy prints of professionally drawn figures may be required if manuscript is accepted.

5. Include a brief abstract (not more than 100 words) summarizing the findings.

6. Submit *five copies* of the manuscript and abstract on white paper. Manuscripts should be addressed to the Editor of the AMERICAN JOURNAL OF SOCIOLOGY, 5835 South Kimbark, Chicago, Illinois 60637. *Manuscripts are not returned to authors.* Enclose a self-addressed stamped postcard to obtain acknowledgment of manuscript receipt. No other acknowledgment will be sent.

7. Each article submitted to the *AJS* must be accompanied by a check or money order for \$15.00, payable to the University of Chicago Press in U.S. currency or

its equivalent. Papers will not be reviewed until this fee has been paid. The submission fee will be waived for papers solely by student authors. Submissions from students must be accompanied by a faculty member's certification of their student status. Citizens of countries with restrictions on the export of U.S. dollars may request waiver of the submission fee.

Format of References in Text: All references to monographs, articles, and statistical sources are to be identified at the appropriate point in the text by last name of author, year of publication, and pagination where appropriate, all within parentheses. Footnotes are to be used only for substantive observations. Specify subsequent citations of the same source in the same way as the first one; do not use "ibid.," "op. cit.," or "loc. cit."

1. When author's name is in the text: Duncan (1959). When author's name is not in text: (Gouldner 1963).

2. Pagination follows year of publication: (Lipset 1964, pp. 61-65).

3. For more than three authors, use "et al." For institutional authorship, supply minimum identification from the beginning of the complete citation: (U.S. Bureau of the Census 1963, p. 117).

4. With more than one reference to an author in the same year, distinguish them by use of letters (*a*, *b*) attached to the year of publication: (1965*a*).

5. Enclose a series of references with a single pair of parentheses, separated by semicolons: (Duncan 1959; Gouldner 1963; Lipset 1964, pp. 61-65).

6. Format for references to machine-readable data files is

Davis, James Allan. 1978. GSS (General Social Survey). NORC ed. Chicago: NORC. Distributed by Roper Public Opinion Research Center, New Haven, Conn.

Format of References: List all items, including machine-readable data files (MRDF), alphabetically by author (providing the full list of multiple authors) and, within author(s), by year of publication, in an appendix titled "REFERENCES." For examples, see text of articles in this issue.

(Rev. 3/98)

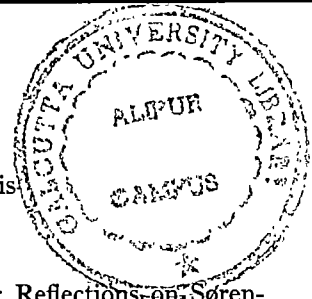
CONTENTS

Symposium on Class Analysis

- 1523 Toward a Sounder Basis for Class Analysis
AAGE B. SØRENSEN
- 1559 Class, Exploitation, and Economic Rents: Reflections on Sørensen's "Sounder Basis".
ERIK OLIN WRIGHT
- 1572 Rent, Class Conflict, and Class Structure: A Commentary on Sørensen
JOHN H. GOLDTHORPE
- 1583 A Neo-Utilitarian Theory of Class?
DIETRICH RUESCHEMEYER AND JAMES MAHONEY

Also

- 1592 State-Society Relations and the Discourses and Activities of the 1989 Beijing Student Movement
DINGXIN ZHAO
- 1633 Embedded Altruism: Blood Collection Regimes and the European Union's Donor Population
KIERAN HEALY
- 1658 The Evolution of Sex Segregation Regimes
MARIKO LIN CHANG
- 1702 Multinomial Logit Latent-Class Regression Models: An Analysis of the Predictors of Gender-Role Attitudes among Japanese Women
KAZUO YAMAGUCHI



Commentary and Debate

- 1741 The Devaluation of Women's Work: A Comment on Tam
PAULA ENGLAND, JOAN M. HERMSEN, AND DAVID A. COTTER
- 1752 Occupational Wage Inequality and Devaluation: A Cautionary
Tale of Measurement Error
TONY TAM

Book Reviews

- 1761 *Department and Discipline: Chicago Sociology at One Hundred* by
Andrew Abbott
ARTHUR L. STINCHCOMBE
- 1763 *Re-Orienting Western Feminisms: Women's Diversity in a Post-
colonial World* by Chilla Bulbeck
LOURDES TORRES
- 1765 *Gender Violence in Africa: African Women's Responses* by Decem-
ber Green
JACQUELINE BHABHA
- 1767 *Asian/American: Historical Crossings of a Racial Frontier* by Da-
vid Palumbo-Liu
YEN LE ESPIRITU
- 1769 *Global Transformations: Politics, Economics and Culture* by David
Held, Anthony McGrew, David Goldblatt, and Jonathan Perraton
GIOVANNI ARRIGHI
- 1770 *The Discipline of Teamwork: Participation and Concertive Control*
by James R. Barker
RANDY HODSON
- 1772 *Taking Care of Business: Samuel Gompers, George Meany, Lane
Kirkland, and the Tragedy of American Labor* by Paul Buhle
KIM SCIPES
- 1774 *Rethinking the Labor Process*, edited by Mark Wardell, Thomas
L. Steiger, and Peter Meiksins
MARTIN PARKER
- 1776 *Dragon in a Three-Piece Suit: The Emergence of Capitalism in
China* by Douglas Guthrie
CHING KWAN LEE

- 1779 *Guests and Aliens* by Saskia Sassen
CHARLES TILLY
- 1780 *Global Futures: Migration, Environment and Globalization*
by Avtar Brah, Mary J. Hickman, and Máirtín Mac an Ghaill
PETER J. TAYLOR
- 1782 *Marketing the American Creed Abroad: Diasporas in the U.S. and Their Homelands* by Yossi Shain
SARAH V. WAYLAND
- 1784 *Ambiguities of Domination: Politics, Rhetoric, and Symbols in Contemporary Syria* by Lisa Wedeen
GEORGE STEINMETZ
- 1786 *Gendering Politics: Women in Israel* by Hanna Herzog
EMILY STOPER
- 1788 *A Generation Divided: The New Left, the New Right, and the 1960s*
by Rebecca E. Klatch
DAVID S. MEYER
- 1790 *Race, Money, and the American Welfare State* by Michael K. Brown
DALTON CONLEY
- 1791 *Being Black, Living in the Red: Race, Wealth, and Social Policy in America* by Dalton Conley
DAVID A. REINGOLD
- 1793 *Something Within: Religion in African-American Political Activism* by Fredrick C. Harris
MARY PATTILLO-MCCOY
- 1795 *Faithful and Fearless: Moving Feminist Protest inside the Church and Military* by Mary Fainsod Katzenstein
MARY BERNSTEIN
- 1797 *Ordaining Women: Culture and Conflict in Religious Organizations* by Mark Chaves
R. STEPHEN WARNER
- 1799 *Congregations in Conflict: Cultural Models of Local Religious Life*
by Penny Edgell Becker
MITCHELL L. STEVENS

- 91 *After Heaven: Spirituality in America since the 1950s* by Robert Wuthnow
PETER DOBKIN HALL
- 1803 *Codependent Forevermore: The Invention of Self in a Twelve Step Group* by Leslie Irvine
DIANE BJORKLUND
- 1805 *The Most Beautiful Girl in the World: Beauty Pageants and National Identity* by Sara Banet-Weiser
MAXINE CRAIG
- 1807 *Teaching America about Sex: Marriage Guides and Sex Manuals from the Late Victorians to Dr. Ruth* by M. E. Melody and Linda M. Peterson
JULIA A. ERICKSEN
- 1808 *Queer in Russia: A Story of Self, Sex, and the Other* by Laurie Essig
ANNICK PRIEUR
- 1810 *A Thousand Screenplays: The French Imagination in a Time of Crisis* by Sabine Chalvon-Demersay
S. CRAIG WATKINS
- 1812 *Mission Impossible: Using Fantasy Documents to Tame Disaster* by Lee Clarke
PETER K. MANNING
- 1814 *Problems of Form*, edited by Dirk Baecker
STEPHAN FUCHS
- 1819 Acknowledgments to Referees
- 1823 Contents of Volume 105
- 1839 Book Reviewers for Volume 105

IN THIS ISSUE

AAGE B. SØRENSEN is professor of sociology and chair of the joint doctoral program in organizational behavior at Harvard University. His research interests are in the areas of social stratification and the sociology of education. The article included in this issue reflects his current efforts to advance the theoretical understanding of class and social status.

ERIK OLIN WRIGHT is professor of sociology and director of the Havens Center at the University of Wisconsin, Madison. Among his interests are class analysis and historical change and political and economic sociology, which are reflected in his work on comparative analyses of class structure, class biology, and class consciousness; Marxism and feminism; foundations of radical social and political theory; and theories of the state.

JOHN H. GOLDTHORPE, an official fellow of sociology at Nuffield College, Oxford University, works on social stratification and comparative macrosociology, including methodological issues. Currently, his research centers on comparative mobility patterns and class structure in capitalist and state socialist societies. He is also interested in the theory of social action, especially rational action theory, and its relation to quantitative sociological analysis.

DIETRICH RUESCHEMEYER is professor of sociology at Brown University. His research interests include the sociology of politics and the state, comparative historical sociology, questions of international development, and social and political theory. With Evelyn Huber and Jon Stephens, he is the author of *Capitalist Development and Democracy* (University of Chicago Press, 1992).

JAMES MAHONEY is assistant professor at Brown University. His research focuses on the comparative-historical analysis of national development, including the study of democracy and authoritarianism, Third World economic growth, and revolution. He is currently working on a book that analyzes liberalism and regime change in five Central American countries during the 19th and 20th centuries.

DINGXIN ZHAO is assistant professor of sociology at the University of Chicago. His interests are in social movements, development, and sociology of mass media. He has spent the last seven years in studying the origins and development of the 1989 Beijing Student Movement. His book *The Power of Tiananmen* will be published by the University of Chicago Press early in 2001.

KIERAN HEALY is a Ph.D. candidate in sociology at Princeton University. His dissertation examines the procurement and exchange of human organs in the United States. His research interests include the sociology of markets, money, and risk, with current focus on the commodification of body parts and products.

MARIKO LIN CHANG is assistant professor of sociology and of social studies at Harvard University. She is currently involved in two lines of re-

search. The first examines variation in patterns of occupational sex segregation in developing countries, and the second addresses the ways that economic attitudes and financial knowledge are associated with the distribution of wealth across race, class, and gender lines.

KAZUO YAMAGUCHI is professor of sociology at the University of Chicago. His principal research interest is to develop new statistical or mathematical models and apply them to research in such areas as social stratification, work and family, rational choice theory, social networks, and drug-use epidemiology.

Information for Contributors

Editorial Procedures: All papers deemed appropriate for *AJS* are sent out anonymously to readers. The editors rely heavily on the judgments of those readers, although they are not bound by them. To protect anonymity, only the title should appear on the manuscript. Attach a cover page with the title of the manuscript, your name, and your affiliation. Manuscripts are accepted for publication subject to the understanding that *AJS* has the right of first publication. Submission of a manuscript to *AJS* is taken to indicate the author's commitment to publish in this *Journal*. No paper known to be under consideration by any other journal will be reviewed by *AJS*. In addition, *AJS* is not a suitable vehicle for materials published elsewhere, in whole or in part. Upon publication, all rights, including subsidiary rights, are owned by the University of Chicago Press. The author retains the right to use his or her article without charge in any book of which he or she is the author or editor after it has appeared in the *Journal*.

Manuscript Acceptance Policy: While it is our policy to require the assignment of copyright on most journal articles and review essays, we do not usually request assignment of copyright for other contributions. Although the copyright to such a contribution may remain with the author, it is understood that, in return for publication, the *Journal* has the nonexclusive right to publish the contribution and the continuing right, without limit, to include the contribution as part of any reprinting of the issue and/or volume of the *Journal* in which the contribution first appeared by any means and in any format, including computer-assisted storage and readout, in which the issue and/or volume may be reproduced by the publisher or by its licensed agencies.

Preparation of Copy

1. Type *all* copy—including indented matter, footnotes, tables, and references—*double spaced*, allowing generous margins at the top, bottom, and sides of page. Please do not break words at ends of lines or justify right-hand margins. Do not use proportional spacing fonts. Indicate italics by underlining only. Number all footnotes, beginning with the acknowledgment note as number 1. *Provide a word count on the cover page of the manuscript.*

2. Begin each table on a separate page. Refer to each table in numerical order in the text. Tables should not contain vertical lines or more than 20 two-digit columns or the equivalent.

3. Clarify all mathematical symbols (e.g., Greek letters) with words in the margins of the manuscript.

4. Original or glossy prints of professionally drawn figures may be required if manuscript is accepted.

5. Include a brief abstract (not more than 100 words) summarizing the findings.

6. Submit *five copies* of the manuscript and abstract on white paper. Manuscripts should be addressed to the Editor of the AMERICAN JOURNAL OF SOCIOLOGY, 5835 South Kimbark, Chicago, Illinois 60637. *Manuscripts are not returned to authors.* Enclose a self-addressed stamped postcard to obtain acknowledgment of manuscript receipt. No other acknowledgment will be sent.

7. Each article submitted to the *AJS* must be accompanied by a check or money order for \$15.00, payable to the University of Chicago Press in U.S. currency or

its equivalent. Papers will not be reviewed until this fee has been paid. The submission fee will be waived for papers solely by student authors. Submissions from students must be accompanied by a faculty member's certification of their student status. Citizens of countries with restrictions on the export of U.S. dollars may request waiver of the submission fee.

Format of References in Text: All references to monographs, articles, and statistical sources are to be identified at the appropriate point in the text by last name of author, year of publication, and pagination where appropriate, all within parentheses. Footnotes are to be used only for substantive observations. Specify subsequent citations of the same source in the same way as the first one; do not use "ibid.," "op. cit.," or "loc. cit."

1. When author's name is in the text: Duncan (1959). When author's name is not in text: (Gouldner 1963).

2. Pagination follows year of publication: (Lipset 1964, pp. 61-65).

3. For more than three authors, use "et al." For institutional authorship, supply minimum identification from the beginning of the complete citation: (U.S. Bureau of the Census 1963, p. 117).

4. With more than one reference to an author in the same year, distinguish them by use of letters (*a*, *b*) attached to the year of publication: (1965*a*).

5. Enclose a series of references with a single pair of parentheses, separated by semicolons: (Duncan 1959; Gouldner 1963; Lipset 1964, pp. 61-65).

6. Format for references to machine-readable data files is

Davis, James Allan. 1978. GSS (General Social Survey). NORC ed. Chicago: NORC. Distributed by Roper Public Opinion Research Center, New Haven, Conn.

Format of References: List all items, including machine-readable data files (MRDF), alphabetically by author (providing the full list of multiple authors) and, within author(s), by year of publication, in an appendix titled "REFERENCES." For examples, see text of articles in this issue.

(Rev. 3/98)

IN THIS ISSUE

HOWARD SCHUMAN is professor and research scientist emeritus at the University of Michigan. He recently published a book with others on *Racial Attitudes in America* and is currently writing a book on collective memory and collective knowledge. Now living on the coast of Maine, he continues his work through the Institute for Social Research in Ann Arbor.

AMY D. CORNING is a doctoral candidate in the department of sociology at the University of Michigan. Her research interests lie in the areas of collective memory, social structural influences on attitudes, Soviet and post-Soviet society, and survey research methods. Her dissertation focuses on Russian immigrants who have settled in the United States over the past two decades, exploring how they remember their history in both political and personal terms.

ERIK OLIN WRIGHT is Vilas Professor of Sociology and director of the Havens Center for the Study of Social Structure and Social Change at the University of Wisconsin, Madison. His work has mainly been concerned with the development of the Marxist tradition in sociology, with particular focus on the analysis of class. His most recent book is *Class Counts: Comparative Studies in Class Analysis* (Cambridge University Press, 1997).

MARSHALL GANZ is an instructor of public policy at the John F. Kennedy School of Government and is a doctoral candidate in sociology at Harvard University. For 32 years, he served as a civil rights, union, electoral, and community organizer.

DANIEL M. CRESS is assistant professor of sociology at the University of Colorado at Boulder. His current research interests focus on issues of power and marginality. He is currently investigating institutional and noninstitutional political activity by Native American since 1890, and student rioting on U.S. campuses from 1980 to the present.

DAVID A. SNOW is professor of sociology at the University of Arizona. He has published widely on various aspects of social movements, framing processes in the context of movements, conversion processes, self and identity, ethnographic field methods, and homelessness. He is coauthor of *Down on Their Luck: A Study of Homeless Street People* (with Leon Anderson) and of *Social Movements: Readings on Their Emergence, Mobilization, and Dynamics* (with Doug McAdam). He is currently working on a book on framing processes in the context of politics and social movements.

NOBUYUKI TAKAHASHI is a Ph.D. candidate in the sociology department at the University of Arizona. His areas of interest include social psychology, group processes, and social networks. He is currently interested in the relationship between social networks and the type of comparison that people use in making fairness judgments.

XUEGUANG ZHOU is associate professor of sociology at Duke University.

He conducts research in the areas of social stratification, organizations, and economic sociology. In addition to his work on the evolution of social stratification patterns in urban China, he also works on a study of employment relationships in multinational corporations and a study of the role of social relationships in market transactions.

YANG CAO is a Ph.D. candidate in the department of sociology at Cornell University. He is interested in economic sociology, organizations, stratification, and research methods. His current research focuses on the impact of institutional change on organizations and stratification in contemporary China.

VICTOR G. NEE is Goldwin Smith Professor of Sociology and chair of the department of sociology at Cornell University. His research interests are in economic sociology, ethnic stratification, and the new institutionalism. He is completing a book with Richard Alba comparing the experiences of European immigrants and post-1965 immigrants and a book on market transition and stratification in China.

Information for Contributors

Editorial Procedures: All papers deemed appropriate for *AJS* are sent out anonymously to readers. The editors rely heavily on the judgments of those readers, although they are not bound by them. To protect anonymity, only the title should appear on the manuscript. Attach a cover page with the title of the manuscript, your name, and your affiliation. Manuscripts are accepted for publication subject to the understanding that *AJS* has the right of first publication. Submission of a manuscript to *AJS* is taken to indicate the author's commitment to publish in this *Journal*. No paper known to be under consideration by any other journal will be reviewed by *AJS*. In addition, *AJS* is not a suitable vehicle for materials published elsewhere, in whole or in part. Upon publication, all rights, including subsidiary rights, are owned by the University of Chicago Press. The author retains the right to use his or her article without charge in any book of which he or she is the author or editor after it has appeared in the *Journal*.

Manuscript Acceptance Policy: While it is our policy to require the assignment of copyright on most journal articles and review essays, we do not usually request assignment of copyright for other contributions. Although the copyright to such a contribution may remain with the author, it is understood that, in return for publication, the *Journal* has the nonexclusive right to publish the contribution and the continuing right, without limit, to include the contribution as part of any reprinting of the issue and/or volume of the *Journal* in which the contribution first appeared by any means and in any format, including computer-assisted storage and readout, in which the issue and/or volume may be reproduced by the publisher or by its licensed agencies.

Preparation of Copy

1. Type *all* copy—including indented matter, footnotes, tables, and references—*double spaced*, allowing generous margins at the top, bottom, and sides of page. Please do not break words at ends of lines or justify right-hand margins. Do not use proportional spacing fonts. Indicate italics by underlining only. Number all footnotes, beginning with the acknowledgment note as number 1. *Provide a word count on the cover page of the manuscript.*

2. Begin each table on a separate page. Refer to each table in numerical order in the text. Tables should not contain vertical lines or more than 20 two-digit columns or the equivalent.

3. Clarify all mathematical symbols (e.g., Greek letters) with words in the margins of the manuscript.

4. Original or glossy prints of professionally drawn figures may be required if manuscript is accepted.

5. Include a brief abstract (not more than 100 words) summarizing the findings.

6. Submit *five copies* of the manuscript and abstract on white paper. Manuscripts should be addressed to the Editor of the AMERICAN JOURNAL OF SOCIOLOGY, 5835 South Kimbark, Chicago, Illinois 60637. *Manuscripts are not returned to authors.* Enclose a self-addressed stamped postcard to obtain acknowledgment of manuscript receipt. No other acknowledgment will be sent.

7. Each article submitted to the *AJS* must be accompanied by a check or money order for \$15.00, payable to the University of Chicago Press in U.S. currency or

its equivalent. Papers will not be reviewed until this fee has been paid. The submission fee will be waived for papers solely by student authors. Submissions from students must be accompanied by a faculty member's certification of their student status. Citizens of countries with restrictions on the export of U.S. dollars may request waiver of the submission fee.

Format of References in Text: All references to monographs, articles, and statistical sources are to be identified at the appropriate point in the text by last name of author, year of publication, and pagination where appropriate, all within parentheses. Footnotes are to be used only for substantive observations. Specify subsequent citations of the same source in the same way as the first one; do not use "ibid.," "op. cit.," or "loc. cit."

1. When author's name is in the text: Duncan (1959). When author's name is not in text: (Gouldner 1963).

2. Pagination follows year of publication: (Lipset 1964, pp. 61-65).

3. For more than three authors, use "et al." For institutional authorship, supply minimum identification from the beginning of the complete citation: (U.S. Bureau of the Census 1963, p. 117).

4. With more than one reference to an author in the same year, distinguish them by use of letters (*a*, *b*) attached to the year of publication: (1965*a*).

5. Enclose a series of references with a single pair of parentheses, separated by semicolons: (Duncan 1959; Gouldner 1963; Lipset 1964, pp. 61-65).

6. Format for references to machine-readable data files is

Davis, James Allan. 1978. GSS (General Social Survey). NORC ed. Chicago: NORC. Distributed by Roper Public Opinion Research Center, New Haven, Conn.

Format of References: List all items, including machine-readable data files (MRDF), alphabetically by author (providing the full list of multiple authors) and, within author(s), by year of publication, in an appendix titled "REFERENCES." For examples, see text of articles in this issue.

(Rev. 3/98)

Collective Knowledge of Public Events: The Soviet Era from the Great Purge to Glasnost¹

Howard Schuman and Amy D. Corning
University of Michigan

We explore the knowledge of a probability sample of Russians in 1994 about nine events that occurred within the past 60 years. We consider three competing hypotheses about how knowledge relates to age: (1) adolescence and early adulthood constitute a critical age for acquiring knowledge of public events; (2) the unique content of an event creates age relations; (3) the primary influence on knowledge is a period effect. We also hypothesize that "years of education" has two different meanings in relation to knowledge: one about socialization that promotes state-approved images of the past, and the other about development of a cognitive sophistication that challenges such images. Partial support for each hypothesis is reported. The relation of collective knowledge to collective memory is also considered.

For all of us there is a twilight zone between history and memory; between the past as a generalized record which is open to relatively dispassionate inspection and the past as a remembered part of, or background to, our own life.

(Hobsbawm 1987, p. 3)

We investigate what ordinary Russians know about a number of events that occurred in their country over six decades, from the Great Purge in the late 1930s to the beginning of glasnost ("openness") in the 1980s. Some

¹ For helpful recommendations at various points, we are indebted to Barry Schwartz, Alex Inkeles, Stanley Presser, Misha Tsypkin, Eleanor Singer, Norbert Schwarz, Irwin Weil, Michael Kennedy, William Zimmerman, and Robert Belli, and to several anonymous journal referees. The research grew out of a project supported by the National Institute of Aging (SR 1 AF08951). We also thank InterMedia (Washington, D.C.), the sponsor of the Russian survey, for allowing us to include questions. The data that form the basis for this article will be archived in the Inter-university Consortium for Political and Social Research, Ann Arbor, Michigan. Direct correspondence to either author at the Institute for Social Research, P.O. Box 1248, University of Michigan, Ann Arbor, Michigan 48106. E-mail: hschuman@umich.edu, corninga@umich.edu

of the events reflect positively on Soviet history (e.g., launching the first space satellites), others negatively (e.g., the Great Purge itself), and together they offer some sense of what Russians of different ages and social backgrounds know as they "discuss, debate, and negotiate the past and, through this process, define the future" (Gillis 1994, p. 20). Because our concern is with knowledge that is shared by substantial parts of a population, we refer to it as "collective knowledge," and because the knowledge consists of memories of events (whether directly experienced or learned about from others), we regard it as falling under the general concept of collective memory.

Theoretically, we explore the influence of two primary sources of knowledge of past public events: cohort experience and education. More specifically, we test three competing hypotheses about cohort effects: one that stresses the importance of adolescence and early adulthood as a critical life stage for learning about events; one that focuses on how the unique content of an event can stimulate learning at a particular age; and one that considers period effects on learning that can extend over almost the entire age range of a population experiencing an event, though not those born later. We test the effects of education by distinguishing the direct socializing influence of schooling from the increased cognitive sophistication that is often associated with advanced educational attainment.

Our approach differs from a cultural orientation to collective memory that is based on evidence about past events gleaned from documents, institutions, or pronouncements by elites, and also from a purely psychological approach that assumes fixed developmental stages during which knowledge is acquired. We begin by acknowledging the first of these differences, and then proceed to state hypotheses that include but go beyond developmental assumptions.

TWO WAYS OF THINKING ABOUT COLLECTIVE MEMORY

The Cultural Level

Sociologists interested in collective memory in large and complex societies have usually moved in one of two directions. The more frequent direction is to investigate in depth particular cultural symbols, including use of such symbols by those in positions of power. A valuable study of this type is Schwartz's (1996, 1998) account of how collective memories of Abraham Lincoln were used during World War II by the president and other political and cultural leaders to provide legitimization, orientation, clarification, and inspiration for the war effort. Schwartz recognizes that his evidence comes entirely from statements at the level of elites, but suggests that since "these image-makers were socialized by the communities they endeavored to reach, . . . their depictions reflected as well as shaped their audience's

conception of Lincoln" (1996, p. 912). It is possible, however, that the aspects of Lincoln drawn on were uniquely appealing to the elites themselves, as Converse (1987) indicates when considering the political effects of memories of the Vietnam War. The elites may simply have projected what was especially meaningful to themselves on to the general public (Fields and Schuman 1976; see Rieder [1994] on political speeches generally).

Although Schwartz is sensitive to the issue of how much the "collective memories" vocalized by leaders actually resonate with the public, his primary interest in these articles is in collective symbols as part of culture. Consistent with this theoretical perspective, much of the literature on collective memory is devoted to "publicly available symbols and meaning systems not reducible to what is in people's heads" (Jeffrey Olick, personal communication, 1998; see Swidler and Ardití [1994] for a similar view). This emphasis clearly applies to the massive undertaking by Nora (1996–98) to study French *lieux de mémoire*, as well as to much other recent writing on collective memory, for example, Irwin-Zarecka (1994), Sturkin (1997), and in large part Schudson (1992).²

Collective Memory at the Individual Level

A different direction for studies of collective memory is to focus directly on memories of past events that are shared to a greater or lesser extent by the individuals who constitute a representative sample of a larger population, insofar as this can be done through interviewing or in other ways. Although this approach is less frequently pursued in sociology than is cultural analysis, it is closer in spirit to that conceptualized by Halbwachs ([1950] 1980), the sociologist most responsible for introducing the term "collective memory" (Olick and Robbins 1998). For Halbwachs the term referred to the actual memories shared by members of real groups, such as families—memories that remain alive only so long as the group itself survives. He also included broader groups as carriers of collective memories—in an earlier book he wrote that "the well-linked traditions and remembrances" of the nobility have "for a long time been the chief upholder of collective memory" ([1925] 1992, p. 128)—and he even allowed for the possibility of "events of national import that simultaneously alter the lives of all citizens" (1980, p. 77). Thus it is reasonable to assume that Halbwachs might have included the shared memories of a traumatic national event like the Great Purge as a part of the collective memory of many

² Also related to collective memory at the cultural level is research on museums, historical societies, and other institutions that construe the past (e.g., Barthel 1996).

Russians who lived through that period and perhaps also of the descendants to whom those memories were communicated. Of course, this leaves open the question of how much such sharing there actually was, a question our data will address directly.

Using a national survey, Schuman and Scott (1989) asked a cross-section sample of Americans in 1985 to mention two national or world events or changes over the past 50 years that "seem to you to have been especially important." Who mentioned which events in response to this open-ended question was then studied in relation to cohort, education, and other social background variables. Subsequently Scott and Zac (1993) and Schuman, Akiyama, and Knäuper (1998) employed the same approach, the former using data from Britain and the latter data from Germany and Japan. Other sociologists have started from lists of past events; for example, Roberts and Lang (1985) asked a sample of Woodrow Wilson fellows to check off from a list of 26 events those that they "vividly remembered," and Lang et al. (1993) used a similar procedure with a sample of German journalists.³

Starting from this focus on the reported memories of individuals, two independent streams of research have converged to point to adolescence and early adulthood as a "critical age" or "critical period" in the life course, when events have their greatest impact on memory.⁴ Within sociology, Mannheim's ([1928] 1952) classic essay proposing a broad set of generational effects that originate at that stage in the life course has been drawn on by a number of studies of memory for public events. The research noted above by Schuman and his colleagues, by the Langs, and by Jennings (1996) all provide support for a specification of Mannheim's ideas in terms of memories, though the Langs are more qualified in their conclusions than are the others.⁵ This research can also be seen as offering a

³ Seldom attempted but clearly desirable is research that traces systematically how symbols seen to be long-lasting parts of culture grow out of and are sustained by individual memories and how individual memories are initially and continuously shaped by cultural symbols. The theoretical literature on the new institutionalism (Brinton and Nee 1998) calls for making these connections, and Greif (1998) presents a relevant example within that context. Such research is probably most easily done as part of ethnographic investigations of relatively homogeneous communities, for example, small mining towns with a history of conflict-filled strikes (Fentress and Wickham 1992), where the focus can be on the emergence of a more or less unitary collective memory out of negotiations among many individual memories (Zelizer 1995).

⁴ We will use the term "critical age," rather than "critical period," in order to minimize confusion with the term "period effects," which is also important to our analysis.

⁵ In addition to using quite specialized samples, the checklist approach taken by Kurt Lang, Gladys Lang, and their colleagues (Lang and Lang 1978, 1990; Lang et al. 1993) may be somewhat less effective in identifying memories than open-ended questioning that requires respondents to produce the memories themselves. See Schuman and

solution to what Nora (1996, p. 505) regards as the "insoluble . . . uncertainties" that arise in attempting to identify a concrete "generation," for a generation can be demarcated by identifying the distinctive memories its members share based on what happened during their adolescence and early adulthood. One starts with memories and works backwards rather than forward from generations, as was the traditional approach taken by Mannheim and his predecessors (Jaeger [1977] 1985).

At about the same time that these developments were occurring in sociology, several cognitive psychologists who were carrying out laboratory experiments on memory discovered a quite similar phenomenon that they variously called the "reminiscence bump," "peak," or "surge"—that is, memories of personal life events tended to go back to adolescence and early adulthood. This work was not stimulated by knowledge of Mannheim or of other sociological writing, but began as a series of studies of autobiographical memories produced in response to word associations. The first investigators—Fitzgerald and Lawrence (1984) and Rubin, Wetzler, and Nebes (1986)—expected to find monotonic relations pointing to recency of occurrence as the prime determinant of memories. They were surprised to discover that, apart from quite recent memories (essentially the most recent year), the teens and twenties constituted the period in life that yielded the largest number of memories. Subsequently these results have been shown to be replicable with Japanese subjects and to include memories of public as well as personal events. In addition, much the same "critical age" emerged for other forms of memory. For example, Larson (1996) in a small study finds that books mentioned as having provided a particularly memorable reading experience produce a somewhat similar relation to age, with the twenties and thirties identified as modal. Mackavey, Malley, and Stewart (1991) analyzed chapter-long autobiographies of 49 eminent psychologists for autobiographically consequential experiences and found that events from ages 18 to 35 had many more explicit reports of emotion than did references to surrounding periods. Pennebaker and Banasik (1997) have connected the timing of later monument construction to the critical age of cohorts when the events commemorated first occurred. Rubin, Rahhal, and Poon (1998) review a large number of studies that lead them to conclude that "adolescence and early adulthood are special times for memory encoding" (p. 3).

Both psychologists and sociologists have recently added objectively assessed knowledge of the past as another form of memory that may be strongest for events that occur during adolescence and early adulthood. Using convenience samples, Rubin, Rahhal, and Poon (1998) report that

Scott (1987) for a comparison of open- and closed-ended approaches to obtaining memories.

knowledge of the Academy Awards, the World Series, and various current events is more accurate when the critical age is involved than when events are drawn from other periods of life. At the same time, Jennings (1996) and Schuman, Belli, and Bischooping (1997) have begun to explore knowledge in probability samples of Americans.

It is this last type of investigation, collective knowledge of public events, that we develop further, both theoretically and empirically, by applying it to a large national sample from the Russian Federation in 1994. Some 2,400 respondents were asked to identify in their own words a set of events that we presented to them. The events had occurred during the half century preceding the dissolution of the Soviet Union at the end of 1991, which transformed the political and social world of all Russians. Our conclusions from the investigation turn out to be substantially different than we initially expected: to contradict T. S. Eliot (1943), our end was not in our beginning.

HYPOTHESES

We test three primary hypotheses about age-related effects on knowledge. For any particular effect, ideally only one hypothesis should account best for the data, but where there appear to be quite different effects, more than a single process may have occurred. Thus the three hypotheses are competitive but not mutually exclusive. We also test one general hypothesis about the effects of education on knowledge. Throughout, the knowledge we deal with is of past public events.

Hypothesis 1: The Importance of a Critical Age for Collective Knowledge

Our data provide a further test of what might be called Mannheim's (1952) "strong hypothesis" about generational effects, namely, that "youth," defined as adolescence and early adulthood, constitutes the "critical age" during which public events are likely to have their greatest impact and therefore to be remembered best in later years.⁶ Mannheim relied in his reasoning on what would now be called "primacy effects": the first encounter with the larger political world defines what is important for naive individuals, whereas later events are assimilated to experiences from the past and are therefore not likely to have as great an impact (see also Ryder 1965). Furthermore, Mannheim believed that first encounters

⁶ "Other things equal" must always be added, since some knowledge of major events like World War II is likely to be obtained by almost everyone. In most cases, our hypotheses are tested within events, not across events.

with national and world events most often occur during adolescence and early adulthood as individuals begin to experience the larger world beyond their own family and neighborhood. His interest was primarily in political beliefs and values, but there is one passage directly relevant to our concern with knowledge: "I only possess those 'memories' that I have created for myself, only that 'knowledge' I have personally gained in real situations. This is the only sort of knowledge which really 'sticks' and it alone has real binding power" (Mannheim 1952, p. 296).

We add two qualifications to the typically general statements about the importance of a critical age. First, it seems likely that the duration of an event will determine how clear a cohort effect we find.⁷ Events can be discrete, like shooting stars, appearing suddenly to the eye of the public, then vanishing from public attention after a short period. Such events should produce relatively clear-cut cohort effects, since their impact should be mainly on an especially impressionable age range among those alive at the time and should be least known to cohorts that appear after the event is over. Less distinct cohort effects should occur for events that extend over many years and affect successive cohorts; such effects will therefore be difficult to distinguish from period effects. Less visible cohort effects might also occur for an event that stimulates continuous attention through one form or another of public commemoration—"the evaluative aspect of chronicling," as Schwartz puts it (1982, p. 377)—and also for an event that resonates in the public mind with other earlier or later events, as in the series of "-gate" scandals that draw their connotative effect from the original incidents of "Watergate" (Schudson 1989; 1992). (Four of our events have a conceptual and historical association with the Great Purge of the 1930s and we will bear this in mind when they are considered.) Yet even in these long-lasting cases, the precise information available about the event may shift over time, and therefore the content of knowledge may vary in subtle ways that yield evidence of a cohort-specific effect. Wherever possible we have tried to examine more detailed elements of event content.

Second, we need to clarify the span of the supposed critical age and its relation to the nature of events. Mannheim (1952, p. 300) conceived of the critical age as ranging from about 17 to 25, but since this was based on his subjective impressions of Germany in the early 20th century, few would expect exactly the same boundaries to hold in other countries or at other times. Schuman and Scott (1981, p. 377) refer to the "teens or

⁷ English and other European languages we have checked have not developed distinct words for events varying in duration. Pearl Harbor, which happened in a single day, is called an event, but so is World War II termed an event, though it extended over several years and consisted of many separate events like Pearl Harbor.

early twenties," Rubin et al. (1998) extend the range to 10–30, Lang and Lang (1978) speak of the "formative years," and Nora (1996) cites still other demarcations. Recently Holmes and Conway (1999) have drawn on Erikson's (1950) discussion of developmental stages to predict that memories for *public* events should be concentrated in the period when respondents were ages 10–19 (roughly the stage of identity formation) and that memories for *personal* events should refer primarily to the time when respondents were ages 20–29 years (Erikson's stage of intimacy vs. isolation). However, it is important to recognize that some public events have a simple and dramatic nature that can be appreciated even by small children (e.g., Schwartz's [1999] account of his experience during World War II). Other events are abstract and complex and probably meaningful only to individuals beyond adolescence. Thus the concept of a critical age may need to be amended to recognize that different events are likely to interest different ages and therefore to show somewhat different cohort effects at a later point in time.⁸ Indeed, for some of the events that we included, learning may even fall outside the critical age as ordinarily defined, and this leads to our second hypothesis, a more radical departure from an exclusive focus on adolescence and early adulthood as a critical developmental stage.

Hypothesis 2: The Importance of Event Content for Collective Knowledge

Much of the literature on cohort effects assumes implicitly that all individuals in a particular cohort will react in much the same way to a public event, and thus treats as more or less random error the considerable number of people who do not fit the modal pattern. Yet a convincing example of specificity within a critical age is reported by Firebaugh and Chen (1995): clear cohort effects on voting occurred for women as a function of the Nineteenth Amendment, but little or no effect was evident for men. Although a gender difference in cohort effects may not seem remarkable

⁸ Research soon after President Kennedy's assassination suggested that quite young children at the time could be affected (Sigel 1965), and Lang and Lang (1983) provide evidence that the same was true to some extent for the events making up Watergate. Whether such immediate involvement at a young age is reflected in later memory is less certain. Schuman and Scott (1989) found that memories of the Kennedy assassination reported by those eight to 12 years old in 1963 were considerably less frequent than reports by those who had been in their teens and even later twenties at the time. Consistent with this result, Greenstein (1965, p. 1) emphasizes the great change that takes place in political awareness between ages nine and 13. More generally, the extensive literature on political socialization deals more often with party identification than with memories for political events, though recently Sears and Valentino (1997) have studied the effects on preadult learning of events that occurred during a particular presidential campaign.

in this case, which is also not one specifically involving memory, it challenges the notion of indiscriminating impressionability regarding national events and points up the importance of event content to particular age-gender interactions and perhaps to other combinations of variables as well.

Even more directly to the point, Schuman et al. (1997) found that men who had been in their early twenties during the late 1960s showed significantly greater knowledge of the 1968 Tet Offensive than did *either* women or older men, and Jennings (1996) reports a similar Vietnam-related finding. A quite plausible interpretation is that in the late 1960s, young men were likely to pay closer attention to events in Vietnam than were either women or older men, because it was young men who faced the prospect of being drafted to fight in that war. In support of this interpretation, knowledge about the 1969 Woodstock rock music festival, an event which did *not* differ in implications for young men and young women, showed a similar cohort effect to Tet but *not* a cohort-gender interaction.

To state the second hypothesis in a competitive form, what appears to be a cohort effect based on a critical age may be due not to a generalized openness at a particular developmental stage, but to an age-related receptivity to the unique content of a particular event. Further, such effects may occur differentially within a cohort, with greater impact on one gender or on one educational level or on those characterized by some other background attribute.

Hypothesis 3: The Importance of Period Effects for Collective Knowledge

The main competitor to *either* type of cohort effect discussed thus far is the possibility that the impact of some events may be so wide that no single cohort, defined in terms of a small number of birth years, develops appreciably greater knowledge than another. In that case, we could find a period effect on all or almost all cohorts alive at the time of the event: they all would be more knowledgeable than cohorts born after the event had ended. Indeed, the term "generation" was once used in this sense, as when Jefferson and other American Founders spoke of the responsibilities of the "living generation" (Peterson 1977). Such a broad period effect should occur for an event that is so unusual as to be novel and of equal interest to just about all ages. This may account for why, despite considerable effort, Schuman et al. (1997) could not locate a cohort effect for knowledge of the events falling under the rubric of "Watergate."

Can this reasoning be pushed still further to suggest that for an event well documented on film or videotape (e.g., the Kennedy assassination in 1963), even those born well after the event has ended are likely to become as knowledgeable about it as those alive when it happened? "Experienc-

ing” a public event, as Zerubavel (1997, pp. 89–92) notes, seldom means direct personal involvement, and most Americans who were adults at the time of the Kennedy assassination experienced it only through television and other media reports, not so very different from those who learn about it through dramatic media documentaries today. Yet even where the medium of learning is the same (notably, television), contemporaneous and later experiences do differ importantly in that the outcome of an event is uncertain in the contemporaneous case but not in the historical case, with greater emotional intensity likely in the former as well. When President Kennedy was killed there were immediate pressing questions about both the cause of the assassination and its consequences for governing. “Especially in the face of a catastrophe, there is an urge to surrender to the most extreme foreshadowing imaginable” (Bernstein 1994, p. 9). But a replay of the same events at a later point, no matter how vividly presented, does not raise the same urgent questions and thus is not likely to have the same long-term impact on knowledge. In addition, an ongoing event stimulates conversations with others (“rehearsals”) that also enhance memory (Brown and Kulik 1977). Thus we hypothesize that for events that have captured public attention broadly, those alive at the time will have acquired and will retain greater knowledge than those who came afterwards, no matter how much the latter are exposed to the history of the event.

Hypothesis 4: Educational Attainment and Collective Knowledge

Education is the most important factor that acts to limit pure generational effects of all kinds, since it allows ordinary citizens to learn about events that occurred not only before their adolescence but even before they were born. It is thus hardly necessary to hypothesize that knowledge of the past will be partly a function of education, for that is so well established that it needs no replication (see Delli Carpini and Keeter [1996] for recent documentation). For this reason and because education is inversely related to age in our Russian data ($r = .41$; $N = 2421$), we include years of education as a control throughout our analysis. However, we also propose that “years of education” can have at least two different meanings, especially for the Soviet era, and that these two meanings point to different levels of association with different kinds of events.

For events that were simple, favorable to the reputation of the Soviet regime, and taught in a uniform way through the schools and the mass media, we hypothesize relatively *low* correlations between years of schooling and knowledge. Most people will have obtained basic information about the event by an early age, and therefore more years of media or school exposure will not add much beyond that. However, knowledge of events unfavorable to the regime and excluded from discussion in state-

controlled media and schools will have been most available to those with advanced education. In the Soviet Union, much more than the United States, access to alternative sources of ideas was a privilege that often accompanied higher education and associated occupations. Furthermore, well-educated individuals learned to read between the lines of state-controlled pronouncements and to seek unofficial (including underground) sources of information (Inkeles and Bauer 1959; Zimmerman 1987). Hence years of education should be related more strongly to knowledge of topics that were unfavorable to the regime and were suppressed than to those topics that were openly taught.⁹ More generally, these opposite predictions about the role of education imply a negative correlation between the extent to which events are known by the population as a whole and the size of the associations of knowledge with years of schooling.

For our study we also obtained two other social background variables expected to have some association with knowledge: gender and urban/rural location. Location is potentially important because in Russia substantial differences in access to information exist between persons in major cities and those in the countryside, and gender has often been shown in other studies to be related to knowledge (Delli Carpini and Keeter 1996). We include both variables as controls throughout and in addition examine the interactions of each with age. Finally, we draw, at important points, on evidence from Soviet magazines and newspapers, although there is not available for Russia the kind of documentation of past radio and television content that exists for the United States.

METHOD OF STUDY

Our data come from a 1994 survey carried out in European Russia (Russia west of the Urals), which comprises nearly three-quarters of the total Russian population.¹⁰ The knowledge of this cross-section sample ($N = 2,421$) was explored by asking in Russian about 11 past events:

⁹ In the United States education is no doubt also more highly related to knowledge of esoteric than to simple events, but the Soviet situation was more extreme, since many events were greatly distorted or concealed entirely from the general public. For example, when the United States won the hard-fought competition to be first in landing a man on the moon, there was only a brief mention of the American achievement on the Soviet nightly news, following such items as a salute to Soviet metalworkers. Access to direct televised coverage of the moon landing, which was available in many parts of the world, was limited in the Soviet Union to the party, military, and scientific elite (Bill Keller, *New York Times Magazine*, June 27, 1999, p. 61).

¹⁰ Five-sixths of the sample is Russian by nationality, so for simplicity we refer to "Russians" throughout, although the sample was designed to represent the entire population and includes other nationality groups. We found no important effects of ethnicity/nationality on knowledge, and therefore do not include it as a variable in our analyses, except for omitting Jews when dealing with anti-Semitism.

This next section concerns a few words and names from the past that come up now and then, but that many people have forgotten. Could you tell me which ones you have heard of at all, and, if you have, what they refer to in just a few words?

The events were read one at a time, and the respondent's answer to each was recorded verbatim so far as possible. The events are shown in chronological order in table 1, with the date listed being the initial point each was available to public attention, though they differ greatly in the degree to which they were salient over a number of years. We had pretested a total of 34 events, chosen with several criteria in mind: to sample events that were both positive and negative in implications for the Soviet regime, both political and cultural, well distributed over time, and that might have appealed to Russians with different age and other demographic characteristics. Then, after pretesting, and constrained by the space available to us in the larger survey, we eliminated those events known to too few or too many respondents to provide sufficient variation for analysis or that presented other problems for interviewing.¹¹ Any selection of events must be somewhat arbitrary, but our choices were made in an effort to sample widely from the Soviet past and of course without knowledge of eventual results. Nine of the 11 events are considered in detail in this article.¹²

¹¹ A memo providing a description of the pretesting and the reasons for event selections can be obtained by writing either of the authors. The 23 events pretested but not included in the final survey were: Collectivization, Valerii Chkalov, Sergei Kirov, German-Soviet Pact, Katyn Massacre, Battle of Stalingrad, Lavrentii Beria, *Sputnik 1*, Yury Gagarin, Twenty-second CPSU Congress, Valentina Tereshkova, Vladimir Vysotskii, Sinyavskii-Daniel trial, Baikal-Amur Mainline, *Soyuz-Apollo*, Afghan veterans, Andrei Sakharov, Chernobyl, *It Isn't Easy to Be Young*, Nina Andreeva, Vilnius TV Center, GKChP (State Committee on the State of Emergency during the 1991 coup), Belovezhskaia Pushcha Accord. (Twelve other events were considered but not pretested because they duplicated in name or date an event already on the list.)

¹² We included in the final survey the names of two Soviet films from the later 1980s—*Repentance* and *Little Vera*—but omit them from this article, first because they are products of the glasnost era and second in order to reduce the complexity of our presentation. *Little Vera*, shown first in 1988, is a film about adolescent rebellion, with sexual content unusually explicit by Soviet standards. It is known especially to younger Russians. The 1987 film *Repentance* (a complex allegorical critique of the Stalin era) shows somewhat more knowledge by middle-aged Russians than by other ages, consistent with an interpretation in terms of its complexity and its resonance with earlier events. Remnick (1993) devotes five pages to *Repentance* because of its political importance as a sign of increasing openness about Stalinism, but only a passing sentence to *Little Vera*. Yet *Little Vera* is known to three-quarters of our Russian sample, *Repentance* to only one-quarter. This does not mean that Russians will mention events such as *Little Vera* when the question calls for using "importance" as a touchstone (see Rieger 1995), but knowledge qua knowledge is a precondition for other forms of memory, since respondents cannot recall as important an event they know nothing about.

TABLE 1
EVENTS FROM THE SOVIET PAST: DISTRIBUTION OF KNOWLEDGE SCORES*

EVENT	DATE	KNOWLEDGE SCORES				EXAMPLE OF SCORE OF 2	WEIGHTED SCORE†
		0	1	2	Total		
Yezhovshchina	1937	49	45	6	100	"In 1937 they arrested and shot innocent people."	57
Doctors' Plot	1953	69	20	11	100	"That was the arrest of doctors just before Stalin's death."	42
Virgin Lands Campaign	1954	22	72	6	100	"The campaign was about opening up the land in Kazakhstan."	84
Twentieth Congress of the CPSU	1956	71	11	18	100	"Stalin's cult of personality was exposed at that congress."	47
Laika	1957	52	6	42	100	"It was the first dog in space, before Gagarin's flight."	90
Cuban Missile Crisis	1962	78	9	13	100	"In Cuba, the Khrushchev-Kennedy clash—we in-stalled missiles there."	35
<i>One Day in the Life of Ivan Denisovich</i>	1962	82	5	13	100	"The book describes life in the gulag."	31
Prague Spring	1968	76	17	7	100	"They put down Dubcek's movement."	31
Katya Lycheva	1986	76	17	6	100	"A Young Pioneer girl was in the USA on a mission of peace."	29

* Based on respondents ages 18 and older, weighted by age, education, sex, and urban/rural location (Moscow, St. Petersburg, other urban areas, rural areas) to match proportions for the population sampled. The base *N* is 2,421 for each event. Scores are percentages.

† Sum of percentages, each weighted by its scoring category (0, 1, and 2).

Respondent accounts of each event were scored "2" if coded as correct, "1" if partly correct, and "0" if incorrect, as shown in table 1.¹³ In addition, an overall weighted score is provided for each event: the sum of the percentages given each score is weighted by its scoring category (0, 1, and 2). The distinction between correct and partly correct is sometimes difficult to make and is even more difficult to hold constant in meaning across items. However, the distinction of *both* degrees of correctness from incorrect (zero) responses is usually quite clear, and so we also draw on nonzero scores as an overall measure of knowledge.¹⁴ The two indicators of knowledge—weighted scores and nonzero scores—are substantially related ($r = .87$); however, there are some noticeable differences for particular items, and we take account of both measures when considering variations in knowledge of events.

Sample and Coding

The survey that included our knowledge questions was carried out by the Moscow-based organization Russian Public Opinion and Market Research (ROMIR), using face-to-face interviews with a multistage stratified probability sample of European Russia, ages 18 and older.¹⁵ ROMIR employed oversampling of Moscow and St. Petersburg, but we use weighting in table 1 to reflect the cities' correct proportions in the total population. In all other analysis, unweighted data are used but a control for location is always included.¹⁶

¹³ Our scoring was factual, not evaluative. Two respondents received the same score if they gave similar accounts of an event, even though one used positive language and one negative language. A memo describing our coding rules in detail can be obtained by writing either of the authors.

¹⁴ Zero scores consist largely of don't know (DK) responses—more than 80% for seven of the events, 76% for the Twentieth Party Congress, and 66% for the Virgin Lands Campaign. The correlation between zero scores and their DK component considered alone is .97 over the nine events, and thus when referring to zero scores we are essentially dealing with DK answers and the coding is unambiguous.

¹⁵ ROMIR also included 15–17-year-olds, but for our analysis we limit the sample to persons age 18 and older, since education is an important variable and the educational levels of the relatively small number of respondents under 18 are seriously misleading; the subsample is also small (70 cases) for separate analysis. Because of this omission, plus the loss of nine cases due to lack of data on education or to other problems, the original ROMIR sample size of 2,500 is reduced to 2,421 for our analysis.

¹⁶ The response rate for the study was 75.8%, and checks were made during the field period to attempt to ensure the integrity of the interviewing, including observation by one of the authors of a small randomly selected set of interviews in both urban and rural areas. Comparison of the survey results for age, gender, education, and urban-rural residence with the 1989 census of Russia shows generally good correspondence, with about 5% underrepresentation of men and a somewhat greater underrepresentation of the least educated—sources of bias common in U.S. surveys also—which we reduce by weighting in table 1 and by including control variables in later

Initial scoring of answers for correctness was done in Moscow by ROMIR with our advice, but a representative subsample of 96 responses to each item was later translated and recoded by one of the authors fluent in Russian. We found the original coding to have been done well and reliably, and changes in scoring that we made for three events had no effect on results. Moreover, in two cases (the Twentieth Congress of the Communist Party of the Soviet Union [CPSU] and the Prague Spring) where we tried a more refined scoring along a five-point scale of correctness, conclusions were essentially the same as for the trichotomous scoring. Therefore, we concentrated our additional coding on other aspects of knowledge about the events, and these codes will be introduced to test specific hypotheses as we proceed. Each of the additional codes had agreement between two independent coders of at least 85%.

Analysis

We carried out a multiple classification analysis (MCA) of the nine knowledge scores (0–2) for each of the events, with birth cohort, education, gender, and urban/rural location as predictors. These adjusted knowledge scores are plotted in most of the figures we present.¹⁷ Summary MCA coefficients are given in table 2 and will be referred to at later points. In addition, ordinary least squares (OLS) regression was used to obtain direct

multivariate analysis. For further details of sampling, see the technical report prepared by ROMIR (1994).

¹⁷ Multiple classification analysis (Andrews et al. 1973), a convenient form of dummy variable analysis, allows easier viewing of means adjusted for other predictors and is valuable when nonlinearity may be important. We use simple random sampling (SRS) statistics, despite clustering and stratification in the sample design, since our conclusions are seldom of a borderline nature that would be changed in important ways by tests for complex sampling, especially when the focus is on multivariate analysis rather than on univariate description (Skinner, Holt, and Smith 1989). Birth cohort in our analysis has 15 ordered categories of four years each, as shown in fig. 1. A four-year span was the smallest that allowed adequate category base *N*'s, which we regarded as above or close to 100. (We stretched the oldest cohort over a wider span both to obtain sufficient cases and to include all remaining respondents, and for this and other reasons associated with advanced age, results for the oldest cohort should be treated with caution. Its median birth year is 1916, which serves as an appropriate reference point when discussing it.) Our measure of education has eight categories, ranging from primary to complete higher, following the Russian system of schooling. (In Russia, as elsewhere, education is correlated with income— $r = .37$ in our study—but adding income to our regressions has only a slight effect on associations, and income itself has small and sometimes nonsignificant relations to knowledge.) Urban/rural location has four categories (Moscow, St. Petersburg, other urban, rural). We tested all interactions between our 15-category cohort variable and each of the three controls, and we report the five instances (out of 27 possible ones) that are significant, usually at well beyond $P < .05$, though only two appear to be theoretically informative.

TABLE 2
ASSOCIATION OF KNOWLEDGE SCORES WITH SOCIAL BACKGROUND PREDICTORS*

EVENT	BIRTH COHORT		EDUCATION		GENDER		URBAN/RURAL LOCATION	
	η	β	η	β	η	β	η	β
Yezhovshchina18	.24	.30	.30	.16	.17	.25	.18
Doctors' Plot14	.22	.40	.41	.17	.17	.27	.17
Virgin Lands Campaign17	.22	.18	.21	.06	.06	.17	.13
Twentieth Congress of the CPSU14	.19	.34	.35	.13	.13	.18	.10
Laika21	.18	.22	.19	.11	.10	.20	.16
Cuban Missile Crisis14	.18	.36	.36	.28	.27	.16	.08
<i>One Day in the Life of Ivan Denisovich</i>15	.21	.44	.45	.12	.10	.25	.14
Prague Spring14	.14	.31	.29	.22	.21	.18	.11
Katya Lycheva18	.15	.24	.21	.03	.04	.14	.10

* Betas are based on multiple classification analysis of the knowledge scores for each event, with all four social background variables included as predictors. See text n. 17 for definitions of the four predictors.

estimates of linear and curvilinear trends for the same knowledge scores, with the same controls, and these results are shown in all figures. Furthermore, we dichotomized the knowledge scores into zero versus nonzero (1 and 2 combined) and used logistic regression to repeat the OLS regressions, thus testing the robustness of *both* the scoring and the original regressions. In no case did this lead to a nontrivial difference in significance level or to any reason to alter a specific conclusion.

Throughout, we treat $P < .05$ as the nominal probability required for considering an effect statistically significant, and in most cases we report more exact probabilities. (The only exceptions are two large but nonsignificant correlations based on an N of only nine for the events themselves.) Finally, although effects due to birth cohort and effects due to aging cannot be formally distinguished, we assume that "age" has the initial role of facilitating encoding of an event, but that once encoded and carried forward, it can be conceptualized as either a cohort or a period effect. This assumption is supported by theoretically formulated predictions of nonlinear relations between birth cohort and knowledge (cf. Firebaugh and Chen 1995, p. 976).

RESULTS

After a brief overview of results, we consider each event in relation to the hypotheses about knowledge outlined earlier. Beginning with events that appear to show the clearest cohort effects, we move toward a reconceptualization that emphasizes, first, the relevance of the content of an event to different ages and therefore different cohorts, and second, broader period effects that occur for most events. We end with a discussion of the role of education in producing official and unofficial forms of knowledge.

Variations in Knowledge

One of the best known events in table 1 (the most nonzero scores and the second highest weighted score) would be entirely unfamiliar to most Americans: the Virgin Lands Campaign. The campaign was initiated in 1954 by Khrushchev, and during its course more than a million Soviet young people were sent to Kazakhstan and other parts of the USSR to expand grain production. Even among Russians ages 18–21 in 1994, a majority was able to give at least a partly correct answer about this agricultural crash program that had waned well before they entered adolescence. On the other hand, the novel *One Day in the Life of Ivan Denisovich* is the least known of all the events in table 1. The publication of Solzhenitsyn's grim account of the gulag became a symbol for both Soviet

intellectuals and Western scholars of liberalization during Khrushchev's Thaw, but four out of five Russians know nothing at all about it.

The fact that the two best known events in table 1 are the Virgin Lands Campaign (using nonzero scores) and (using weighted scores) the dog Laika, who was launched into orbit in *Sputnik 2* in 1957, indicates that we are not dealing mainly with memories resulting from revelations during glasnost. Both events occurred much earlier in time, were positive (Laika) or mixed (the Virgin Lands Campaign) in implications for achievements under Soviet rule, and were not important topics during glasnost.¹⁸ It also seems unlikely that degree of knowledge of these nine events is due to recency of occurrence more generally, since the correlation between dates and scores is strongly in the opposite direction ($-.49$ with weighted scores and $-.52$ with percentage of nonzero scores, neither significant with nine cases). Of course, it is possible that the earlier events were intrinsically more meaningful to respondents than the later ones; we have no way of separating intrinsic personal importance from historical time.

We suspect that more significant than either date or political importance is a combination of the "human interest" content of the event and its celebration by Soviet schools and mass media. Both factors are doubtless reflected in the high identification of Laika, whose 1957 orbit in space provided a moment of wonderment and accomplishment to most Russians at the time. Both factors were also involved in the Virgin Lands Campaign, as we will indicate later. Yet celebration cannot explain another event in table 1 that is relatively well known (second highest in nonzero scores): the Yezhovshchina, or "time of Yezhov," named for Stalin's NKVD (secret police) chief, who presided over the Great Purge during the years 1936–38. Knowledge of that traumatic time of some 60 years earlier has been transmitted broadly through the population.

We now present each of the nine events in terms of its relation to birth year, which we use to connect the knowledge of respondents with the time an event occurred. Knowledge was, of course, measured only in 1994, and its patterning is used to infer who obtained and then retained the knowledge best from earlier years. But even where the knowledge may have been substantially added to during glasnost for more than the youngest cohorts, almost any plausible model that could account for the patterning

¹⁸ We obtained from two well-educated Russians and one American specialist on Russia independent ratings of the amount of discussion during glasnost of each of our nine events. There was very high agreement among these raters, with the Yezhovshchina and the Doctors' Plot at the high end of discussion, Laika, the Virgin Lands Campaign, and Katya Lycheva having little or no discussion, and the remaining events in between. Glasnost can be considered to have extended from 1986 until the Soviet Union itself was dissolved at the end of 1991, though of course discussion of past events did not end at that point.

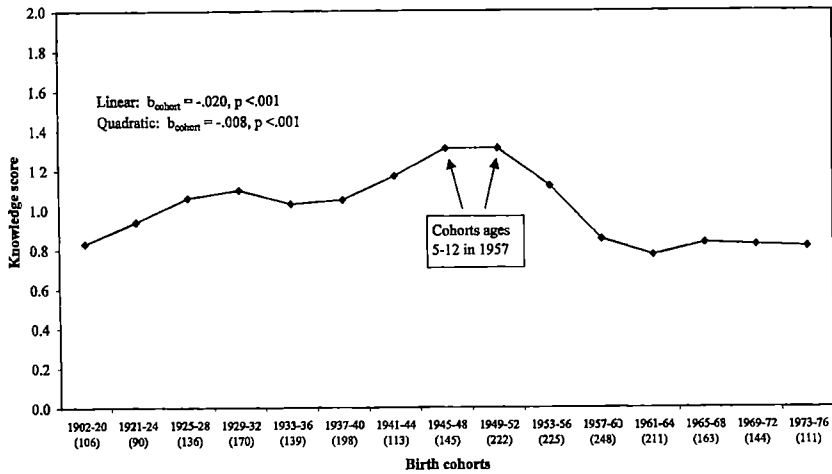


FIG. 1.—Knowledge of Laika (1957) by cohort, controlled for education, gender, and urban/rural location.

requires the assumption of greater “preparation” of older cohorts because of their earlier experience. Thus knowledge in 1994 does point to the time in which it was gained.

TWO EVENTS THAT YIELD DEFINITE COHORT EFFECTS OF UNEXPECTED ORIGIN

Laika and the Importance of Early Childhood

We begin with an event from the mid-1950s that occurred out of the blue, literally as well as figuratively. Laika, the small dog sent aloft as the first mammal in space, had been chosen for fur coloring that would show up well in television transmission and for her good temperament, thus providing an ideal focus of interest at the dawn of the space age. That the effort succeeded is indicated by the high degree of knowledge about Laika across the entire sample more than 35 years after the date of the event. But as figure 1 shows, the most knowledgeable Russians were those born between 1945 and 1952, which means that they were only 5–12 years of age in 1957 when Laika circled the globe in *Sputnik 2*. (The quadratic term in the overall regression is highly significant, and if a narrower range around the peak cohorts is defined, the term becomes still more significant.) A dog traveling across the sky in a sputnik must have made a considerable impression on young children at the time, and thus an event with special appeal to children shows that the age for developing collective

knowledge can be well below adolescence.¹⁹ Older cohorts on the scene at the time were less knowledgeable in 1994, as were younger cohorts that grew up after the event had occurred.

At the same time, if we exclude the cohorts from 1941 to 1956 that show the greatest knowledge, most other Russians alive at the time of Laika's flight demonstrate significantly more knowledge as a group than those born after the event ($P < .001$)—an overall period effect attributable simply to being alive in 1957. Of course, even those born after *Sputnik 2* are somewhat informed about Laika, but their considerable exposure through schools and mass media to the triumphs of the Soviet space program could not match live experience of the event itself.

Katya Lycheva and an Early Age-Gender Interaction

In 1983 a 12-year-old American girl, Samantha Smith, had written Soviet leader Yury Andropov to express support for peaceful coexistence, then traveled to the Soviet Union for the same purpose. In 1986 the Soviet Union produced its own youthful emissary for peace, an 11-year-old girl named Katya Lycheva, who visited America on her mission. Her visit received substantial coverage in the Soviet press and considerable attention in schools, but after her brief national fame she was not heard from again.²⁰

In 1994 we found Katya Lycheva to be about as well known to Russians as the 1968 reform movement in Czechoslovakia (cf. Lycheva with the Prague Spring in table 1). But as figure 2 shows, knowledge of Lycheva is concentrated in the youngest cohorts, those who were 10–17 years of age in 1986 when she came to prominence. (The shape of knowledge for the Prague Spring, which we consider later in fig. 11, is quite different, and specifically it is lowest for the youngest cohort.) Furthermore, eight of the nine events listed in table 2 show men to be significantly more knowledgeable than women, but Katya Lycheva provides the one case where women tend to have higher knowledge scores. More important,

¹⁹ "Childhood amnesia" makes it unlikely that memories of public events can go back much earlier than age 5 (Pillemer 1998). Thus we exclude the first four years of childhood when we consider the effects of experience on knowledge of public events.

²⁰ During Katya's visit to the United States, the teachers' newspaper *Uchitel'skaia gazeta* ran front-page descriptions of her activities in every issue. *Komsomol'skaia pravda*, the newspaper of the communist youth organization, likewise offered detailed coverage of Katya's U.S. travels. Yet just a few months later, when she appeared at the Goodwill Games in Moscow together with Samantha Smith's mother, these same newspapers mentioned her attendance only briefly, despite extensive coverage of other aspects of the games. See, e.g., *Uchitel'skaia gazeta*, March 20 and 22, April 5, and July 8, 1986; *Komsomol'skaia pravda* April 4 and July 6, 1986.

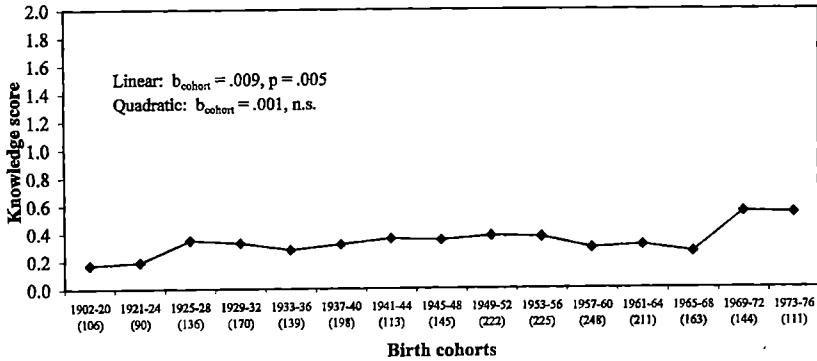


FIG. 2.—Knowledge of Katya Lycheva (1986) by cohort, controlled for education, gender, and urban/rural location.

this gender difference is located almost entirely within the two youngest cohorts, where it is significant ($P < .02$) and produces the specification of cohort effects shown in figure 3. The relation that appeared in figure 2 disappears entirely for men (neither the linear nor quadratic effects for men approach significance, despite the apparent small rise for young men), but the linear relation for women is highly significant, and an additional product term for the overall cohort-by-gender interaction is also significant ($P < .03$).

All respondents in our sample were alive and able to follow the news about Katya Lycheva in 1986, but evidently it was young girls who paid special attention to her and retained that knowledge in their memories

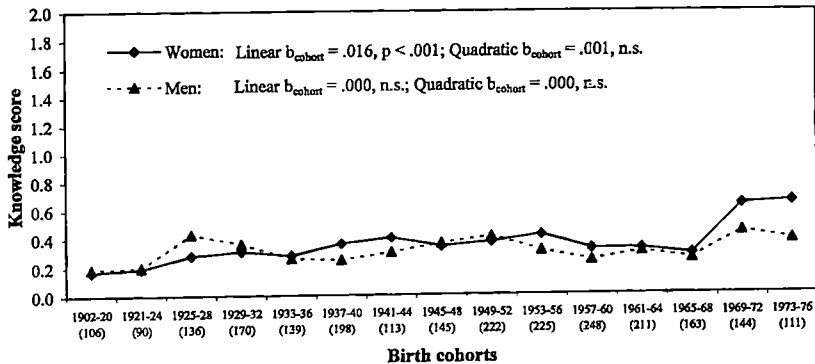


FIG. 3.—Knowledge of Katya Lycheva (1986) by cohort and gender, controlled for education and urban/rural location.

some eight years later when our survey took place. In one sense, this is consistent with the importance of a critical age for gaining knowledge of public events, since women who were even a little older at the time less often knew who Lycheva was. But in another sense, what is demonstrated here is not so much a generalized critical age for knowledge, but rather the appeal of the unique age-gender content of the event to a particular age-gender combination within the total population. We should stress that the singular effect on girls was not because Lycheva was presented to the Soviet public in terms of gender, for the primary emphasis was on her propaganda mission, though her age and gender were of course evident to the public.

Although both Laika and Katya Lycheva indicate an appeal to young people, they seem to do so because of their special content, further limited to girls in the case of Katya. Thus it is not certain how much youth in the sense of a generic critical time of life is involved in either case.

TWO POSSIBLE COHORT EFFECTS

The Cuban Missile Crisis and Mature Adulthood

The "Caribbean Crisis," as it was known in the Soviet Union, brought Soviet-American relations to the brink of nuclear war in 1962. But although the crisis should have made some impression on Russians paying attention at the time, it received much less media emphasis in the Soviet Union than in the United States, and was discussed in only a limited fashion in subsequent Soviet historical accounts.²¹ In addition, despite its international importance, the Cuban Crisis was of brief duration.

The results for the crisis in figure 4 are more difficult to interpret than those for Laika and Katya Lycheva. There is some evidence for a specific cohort effect among Russians who were in their early thirties at the time (the 1929–32 cohort), with the quadratic term for cohorts born between 1925 and 1936 significant ($P < .05$). More generally, those born in 1952 or earlier (i.e., 10 years of age or older during the crisis) are more knowledgeable than those born after 1952 ($P < .001$).²² This can be taken as

²¹ High school textbooks included a brief description of the incident, accusing the United States of having caused the crisis by preparing to invade Cuba (Gardner 1988). Accounts during the 1970s and most of the 1980s presented a similar viewpoint (Garthoff 1989). During glasnost more complete discussion became possible, and in 1989 Soviet, U.S., and Cuban officials who had been involved in the crisis met in Moscow to discuss it more fully (Tolz 1990).

²² The location of the 1949–52 cohort is ambiguous, but the difference between the two levels of knowledge remains significant no matter how that cohort is classified.

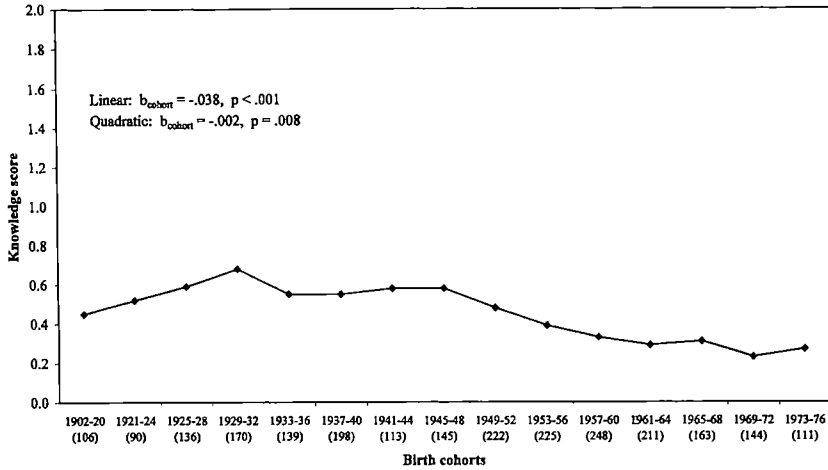


FIG. 4.—Knowledge of Cuban Missile Crisis (1962) by cohort, controlled for education, gender, and urban/rural location.

evidence for a period effect on all those able to attend to the crisis as it happened. In either case, the extension in the opposite direction to that needed for knowledge of the dog Laika makes good sense in terms of event content, because the Cuban Missile Crisis was a distant and abstract occurrence that touched few Russian lives directly. Therefore, it was of primary interest to mature adults, just the opposite of Laika's appeal to young children. Thus the content of an event again becomes an important factor in understanding how its impact varies by age, regardless of whether we focus on the limited cohort effect or on the broader period effect.

One Day in the Life of Ivan Denisovich

Solzhenitsyn's powerful novel about the Soviet gulag was first published in the journal *Novy mir* in 1962, near the end of the thaw associated with Khrushchev's name. It was then banned completely and even ripped out of library copies of *Novy mir* when Khrushchev was deposed in 1964. Knowledge of the novel shows two different age effects in figure 5. On the one hand, there is a broad period effect, such that those born in 1952 or earlier are more knowledgeable than those who came of age after the

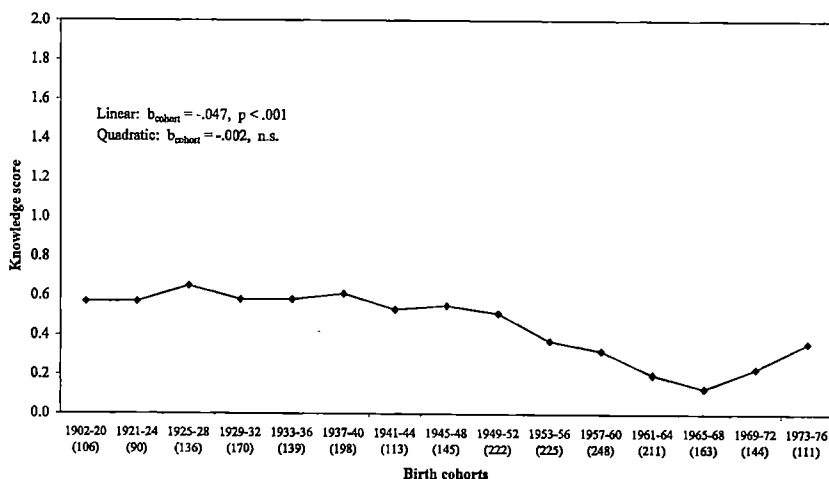


FIG. 5.—Knowledge of *One Day in the Life of Ivan Denisovich* (1962) by cohort, controlled for education, gender, and urban/rural location.

book was banned ($P < .001$).²³ On the other hand, there is a sharp rise in knowledge among the youngest Russians ($P < .03$ for an added quadratic term for the five youngest cohorts, and $P < .001$ if extended to include the seven youngest).²⁴ This upturn is almost certainly due to the publication and publicizing of Solzhenitsyn's writings beginning in 1989 and extending into the 1990s, including even a stage dramatization of *One Day*

²³ The same significance level is found if the 1953–56 cohort is treated as older rather than younger. Within the range from the oldest through the 1952 cohort, those born before 1941 are significantly more knowledgeable ($P < .001$) than those born between 1941 and 1952. This can probably best be attributed to their having experienced the effects of the Great Purge. Although *One Day in the Life of Ivan Denisovich* is set in the postwar years, it is clear from reviews published at the time that readers interpreted it as concerned with the entire Stalinist system of repression. (See, e.g., "Ivan Denisovich, His Friends and Foes," *Novy mir*, January 1964.) We also analyzed further the content of the correct/partially correct responses to *One Day in the Life of Ivan Denisovich* and noted that a small proportion (8%) referred explicitly to the book having been published during the Thaw that followed Khrushchev's Twentieth Party Congress speech. These respondents tended especially to come from the 1941–44 cohort, which was between ages 12 and 15 at the time of Khrushchev's speech and five years older at the time of the novel's publication. The quadratic term for the curve having that cohort as its peak reaches significance in a logistic regression ($P = .023$).

²⁴ The nonsignificant ($P = .066$) quadratic term for the full 15 cohorts is negative, evidently reflecting the overall curve downward from the oldest Russians to the youngest. But when the regression is run using only the youngest cohorts, the quadratic term is positive since the curvilinearity becomes convex.

(Remnick 1993). Since the novel was not addressed to any particular age group, the increase in knowledge by younger Russians seems most interpretable as a critical age effect on those who had passed through adolescence during the extraordinary events of glasnost and the rapid dissolution of the communist state. The slightly older cohorts that had come of age during the last years of the Brezhnev era are the least knowledgeable of all Russians on this point. Thus, overall knowledge of Solzhenitsyn's novel appears to produce two distinct effects: first, a period effect visible among all cohorts alive during the book's publication; second, a critical age effect at a time of great upheaval in the Soviet Union.²⁵ We will consider the theoretical meaning of the unusual critical age effect at later points and especially in our conclusion.

FOUR EVENTS THAT SHOW OTHER SPECIFICATIONS IN TIME

We now turn to four events that were long lasting, though in different ways. For all four, it is difficult to identify a cohort effect in terms of overall knowledge, but we are able to connect more specific event content to the acquisition of knowledge by particular cohorts.

The Virgin Lands Campaign

The most obvious long-lasting event included in our survey was the Virgin Lands Campaign, Khrushchev's much-heralded program to improve grain harvests that sent young volunteers to cultivate land in Kazakhstan. Although the campaign was eventually recognized as ill-conceived—it resulted in widespread soil erosion—the effort continued from its first announcement in 1954 through Khrushchev's years of leadership, and then at lower levels through most of the 18 Brezhnev years as well. The twentieth anniversary of the campaign was celebrated in 1974, by which time over 1 million citizens had received medals "For the Development of the Virgin Lands," and Brezhnev spoke of the history-making feats of the

²⁵ In the case of *One Day in the Life of Ivan Denisovich* there are also statistically significant interactions between our 15-category cohort variable and each of the other background variables: education, gender, and urban/rural location. The interactions appear to occur because in each case those adults in one group (men, the higher educated, those living in Moscow or St. Petersburg) when the book was first published are much more knowledgeable than the other group (women, less educated, rural), but this difference decreases for younger cohorts. The more knowledgeable groups therefore have steeper slopes than the less knowledgeable groups. We should also emphasize that knowledge of Solzhenitsyn's novel is largely concentrated among the best educated and also among men in the two major cities, so that when we speak of its spread across many age groups, this is mostly within a relatively narrow stratum.

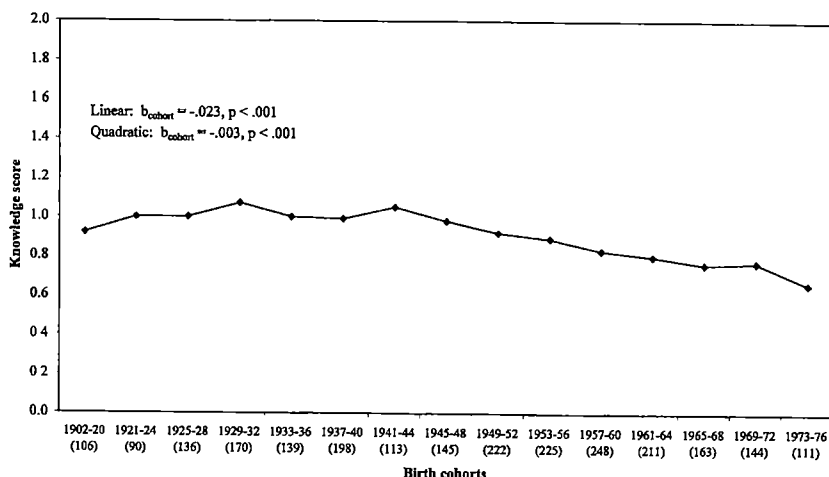


FIG. 6.—Knowledge of Virgin Lands Campaign (1954) by cohort, controlled for education, gender, and urban/rural location.

first settlers as “forever inscribed in the memory of the people.”²⁶ These commemorations were widely reported in Soviet newspapers and the proceedings were published as a book.²⁷ In 1978 Brezhnev published a further book focusing on his own leadership role in the Virgin Lands Campaign (and virtually ignoring Khrushchev’s role), which was widely read and discussed in schools throughout the Soviet Union.²⁸ Thus the Virgin Lands Campaign figured in the news for more than two decades, and even in the 1980s it was never explicitly repudiated.

Figure 6 shows a mild but statistically reliable overall curvilinearity in knowledge, with a somewhat uneven peak. (The slight U-shape between the 1929–32 and 1941–44 cohorts does not approach significance and is probably best regarded as due to sampling error.) The main drop in knowledge occurs for those born toward the end of the 1950s and later, who reached their teenage years after the campaign had trailed off, though even the youngest cohorts are more knowledgeable about the Virgin

²⁶ *The Great Achievement of the Party and the People: Speeches by Participants in the Meeting Held in Alma-Ata on March 15, 1974, to Mark the Twentieth Anniversary of the Opening Up of the Virgin Lands* (Moscow: Novosti, 1974), p. 11.

²⁷ *The Great Achievement of the Party and the People: Materials from the Ceremonial Session in Alma-Ata, 1974* (Moscow: Politizdat, 1974). See newspaper coverage on March 15 and 16, 1974, but also throughout the following and preceding weeks.

²⁸ See, e.g., *Uchitel'skaia gazeta*, November 11, 14, and 30, 1978, where articles document the extensive use of Brezhnev’s book in secondary schools.

Lands Campaign than they are about most other events. Since publicity about it continued over a considerable span of time, it is not surprising to find less cohort specificity in knowledge scores than that which occurs for the events discussed previously.

In order to connect more specific knowledge to points within the nearly three decades that the campaign received public attention, we coded the correct/partially correct responses for mentions of the two Soviet leaders whose names were associated with the campaign in different periods. The 170 respondents who mentioned only Khrushchev's role had a mean age of 50 in 1994, indicating that they were 10 in 1954 when the campaign began and in their adolescence during its early years. The 142 respondents who mentioned only Brezhnev's role had a mean age of 39, indicating that they were age 10 as Brezhnev's rule began and into early adulthood when his book was widely promoted (for the difference in means, $t = 6.3$; $df = 305$; $P < .001$). (Only eight respondents mentioned both leaders, and their mean age in 1994 is, as might be expected, intermediate: 45.) Thus, even though general knowledge of the Virgin Lands Campaign has little specific location among cohorts, an important element of that knowledge can be tied more closely to personal experience during younger years. Russians who were young at the inception of the campaign in the 1950s continued to associate it with Khrushchev's name, despite state and party propaganda throughout later years that presented it as largely Brezhnev's doing. Although this can be treated as a cohort effect, it is important to recognize that the campaign was aimed primarily at Russian youth; therefore an alternative interpretation is that the content of the event was especially relevant to a particular age range, much as was true for Laika and Katya Lycheva.

Yezhovshchina, the "Time of Yezhov"

No event of the last 60 years experienced more vicissitudes in Russian public attention than the Great Purge. The purge itself (1936–38) was widely known at the time because of substantial publicity about the activities of purported spies, "wreckers," and other "enemies of the people," and because of the great numbers arrested and imprisoned or executed.²⁹ It

²⁹ The total number of people arrested during 1937–39 is not known, but estimates vary from about 3 million to about 8 million. Scholars also disagree about the level of fear felt by the Soviet public during the late 1930s, but there is general agreement that Soviet citizens were aware of the arrests and trials—whether or not they believed such actions were justified. The newspapers printed frequent warnings about the threat from "enemies of the people," reported arrests, and even urged schoolchildren to be vigilant against "class enemies." See, e.g., *Pravda*, May 10, 12, and June 12 and 13, 1937. Films like *Great Citizen* helped to publicize and justify the activities of the secret police through their portrayal of evil traitors.

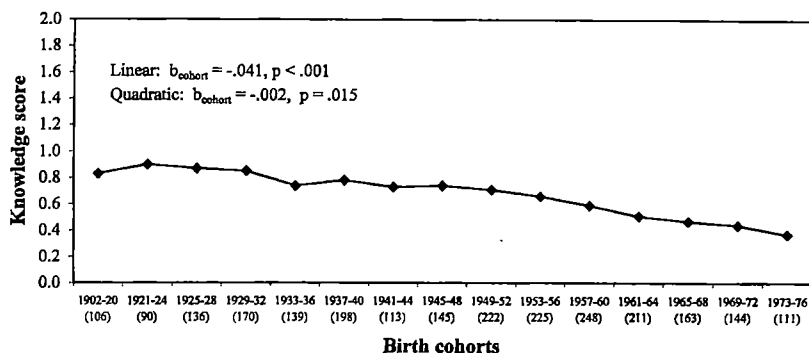


FIG. 7.—Knowledge of the Yezhovshchina (1937) by cohort, controlled for education, gender, and urban/rural location.

was then overshadowed by World War II and the war's immediate aftermath, but briefly resonated again with Stalin's 1953 accusations of a Doctors' Plot (which we deal with below). The Great Purge was brought to light once more in 1956 through Khrushchev's "secret speech" condemning Stalin at the Twentieth Congress of the CPSU and during the subsequent Thaw. It was largely suppressed again for two decades after Khrushchev's ouster in 1964, but finally returned to public view when, beginning in 1987, the crimes perpetrated under Stalin received increasingly open discussion.

Over this long stretch of time, most cohorts had an opportunity to learn at least something about the time of Yezhov, and as noted earlier it turned out to be one of the best known events we asked about. Yet the overall impression conveyed by figure 7 is of greatest knowledge by those who lived through the purge, then a gradual loss of knowledge from that time through cohorts more and more distant from it, with no evidence of a particular increase as a result of glasnost. Neither distant family memories nor very recent revelations could provide for younger cohorts the same degree of knowledge possessed by older Russians who had experienced the Yezhovshchina.

The most visible drop in knowledge appears after those who were born in 1932 or earlier and were thus five years of age or older during the purge.³⁰ If we collapse the four oldest cohort categories and compare them with the next four also collapsed, the difference is highly significant ($P < .001$). Thus Russians who were alive during the time of Yezhov are more

³⁰ Gorbachev (1996) reports having been much affected at age six by the arrest of his grandfather in 1937.

knowledgeable than those born even shortly afterward. Within the four cohorts alive during the purge, however, there is no clear or significant indication of a more limited cohort effect. The Great Purge evidently affected all cohorts alive at the time to much the same extent in terms of later knowledge. The subsequent gradual, rather than precipitous, decline suggests later learning, albeit weaker and weaker over time.

Even though figure 7 shows only a gradual decline over cohorts in the "amount" of general knowledge of the Yezhovshchina, we believed that more subtle differences in knowledge content might reflect the period in which the knowledge was obtained. Khrushchev's secret speech to the Twentieth Party Congress in 1956 was in part a commemoration of the victims of the Great Purge, with statements like the following: "Many thousands of honest and innocent Communists have died as a result of . . . the fact that all kinds of slanderous 'confessions' were accepted, and . . . of the practice of forcing accusations against oneself and others" (Whitney 1963, p. 231).³¹ Khrushchev stopped short, however, of a complete condemnation of Stalin, which would have risked undermining the legitimacy of his own power and of the system as a whole. But by the time of glasnost in the late 1980s, the victims were long since gone and the emphasis was on the larger evil that had been perpetrated in the name of communism over the years and on more general issues of blame and responsibility.

We coded all Yezhovshchina responses that had been scored as correct or partly correct to create three dichotomous variables (each scored "1" for mention of a theme and "0" for lack of mention):

1. *Experience* indicates those who spoke in personal terms about individuals affected (e.g., "My father was arrested"), as against those who made no such reference.
2. *Victims* indicates those who spoke in nonpersonal terms about victims of the purge (e.g., "Many innocent people were killed"), as against no such reference.
3. *Perpetrators* indicates those who referred to the perpetrators of the purge (e.g., "Stalin and Yezhov ordered it"), as against no such reference.

A residual code identified responses not classifiable within this scheme.

Figure 8 shows the use of logistic regression coefficients to graph predicted probabilities by cohort for mentions of each of these three themes

³¹ Although delivered at a closed session of the CPSU congress, knowledge of the speech was quickly transmitted through party organizations across the country.

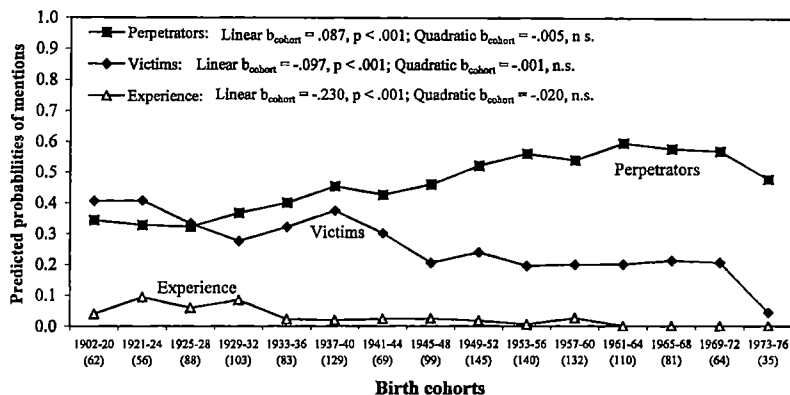


FIG. 8.—Male Muscovites' predicted probabilities of mentioning three components of the Yezhovshchina. (Probabilities are predicted using logistic regression, with N s based on correct/party correct responses.)

in identifications of the Yezhovshchina.³² We naturally anticipated and found mentions of personal experience to be largely restricted to the older end of the cohort continuum, with no substantial drop until after the 1929–32 cohort that was in early childhood (five to eight years old) at the height of the purge. The victims curve is similar in that it shows a general decline from the earliest to the most recent cohorts, but with an interesting exception: an upward bulge for the 1933 to 1944 cohorts, with the peak for ages 16 to 19 at the time of Khrushchev's extraordinary speech to the Twentieth Party Congress. (If we focus on the bulge itself—the four cohorts demarcated by 1929 at one end and 1948 at the other—the quadratic term consistent with curvilinearity is significant at $P < .02$). It appears that an emphasis on victims comes partly from those young when the purge occurred (though our sample lacks Russians who were already old during the purge, so we cannot say how such people would have answered), and then again draws from Russians in their youth at the time of Khrushchev's 1956 speech.

³² A small proportion of respondents (6.3%) mentioned more than one of the three themes. However, making the curves mutually exclusive by omitting overlaps or assigning them arbitrarily to particular categories does not alter the picture, and the overlap is conservative by reducing slightly the differences among the curves. In the figure we plot predicted probabilities of mentioning each of the individual components by cohort, with education controlled, for male Muscovites. Predicted probabilities for the other gender/residence groups follow a similar pattern. (Note that the base N s in fig. 8 are smaller than in other figures, since only correct/party correct responses are included.)

Finally, the perpetrators curve presents a quite different pattern: relatively low for the older cohorts and rising toward the younger cohorts, with maximum mentions by those born in the 1960s and therefore in their late teens and twenties when glasnost led toward the full rediscovery of this history. The tendency of younger people to focus on perpetrators probably indicates their assimilation of the Yezhovshchina to the more general emphasis during glasnost on Stalinist crimes, which were not limited to the 1936–38 purge. Furthermore, the sharp drop in mentions of *any* of the three themes by the youngest cohort—born in 1973–76 and thus coming of age after the end of glasnost and indeed after the end of communist rule—indicates that they tended to give only vague responses like “repressions” that were coded into the residual variable.³³ The focus of this youngest cohort was on the present and the future, not on what to them was the distant past. Thus youthful age does seem to play a distinctive role here in terms of the content of the knowledge absorbed for the Yezhovshchina. This supports the idea suggested in our discussion of *One Day in the Life of Ivan Denisovich*: it is especially during a time of social upheaval that youth—lacking a personally experienced foothold in the past—becomes the part of the population most open to new ways of thinking.

The Twentieth Congress of the CPSU

The Twentieth Congress of the CPSU, at which Khrushchev gave his speech condemning Stalin’s crimes, occurred in 1956. The most visible change in figure 9 occurs for Russians born after 1948, a point that separates those alive and over age eight at the time of the congress from those who came later.³⁴ This difference between two levels of knowledge is highly significant ($P < .001$), and it accounts for most of the overall highly significant linear trend for the entire figure. (If the bridging 1949–52 cohort is omitted, neither of the two parts of the figure taken alone shows a significant linear trend.) We can also see that among Russians alive and beyond early childhood at the time of the speech, those who were 40 were

³³ Responses classified as residual (37% of all nonzero answers) occurred about equally across cohorts, except that their proportion (of correct/partially correct responses) rises sharply for the youngest cohort, indicating that what was transmitted to young Russians in the late 1980s and early 1990s was a general sense of Stalinist repression, without much detail learned or remembered.

³⁴ We do not assume that many eight year olds were paying attention to the Twentieth Party Congress, but reports of Khrushchev’s speech circulated widely following the congress, and by then older children could have learned about it through their families and others. There were some 1,500 delegates to the congress, including prominent foreign communists.

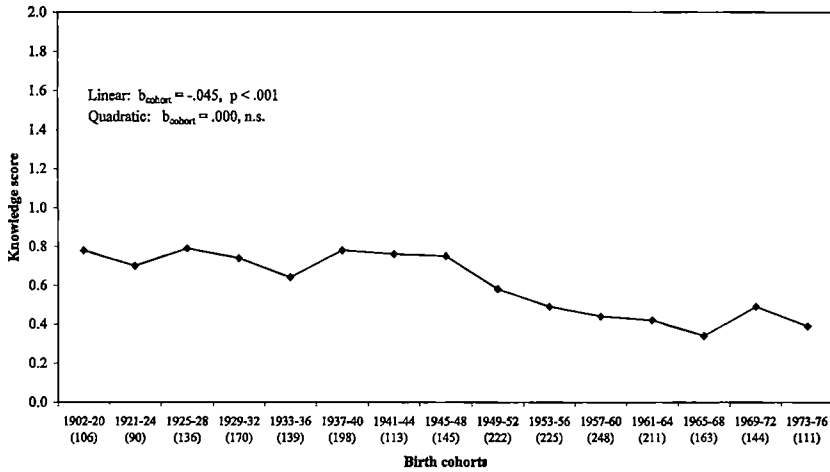


FIG. 9.—Knowledge of the Twentieth Congress of the CPSU (1956) by cohort, controlled for education, gender, and urban/rural location.

as knowledgeable as those who were much younger. Unlike Laika and Katya Lycheva, nothing about the secret speech focused specifically on youth; indeed it clearly drew on memories dating back to the Great Purge. It is simplest to see figure 9 as presenting a period effect on those alive in 1956, with diminished knowledge for all those who came on the scene later. (The slight rise at the younger end of the cohort continuum may reflect information gained during glasnost, but it does not approach significance and without further evidence cannot be distinguished from sampling error.)

It is useful to note that the oldest cohort, born 1902–20, is among the most knowledgeable with regard to the Twentieth Congress. If we take their median birth year of 1916 to locate them in time, then they would have been 40 years old at the point of Khrushchev's speech to the congress, well beyond any definition of early adulthood. Yet by contrast they were relatively low in knowledge of Laika (fig. 1), an event that occurred at almost the same point in time (1957) as the congress (1956), which argues against treating their knowledge of the congress as due simply to a general period effect covering all events. Instead, we interpret their unusually knowledgeable position on the Twentieth Party Congress as drawing on its clear connection to the Great Purge in 1937 when this oldest cohort was about 21 years of age. Following this reasoning, we would also expect—and do find—this same cohort to be among the more knowledgeable about the Solzhenitsyn novel and about the Doctors' Plot, both of which related to—and resonated with—the time of Yezhov. (The only

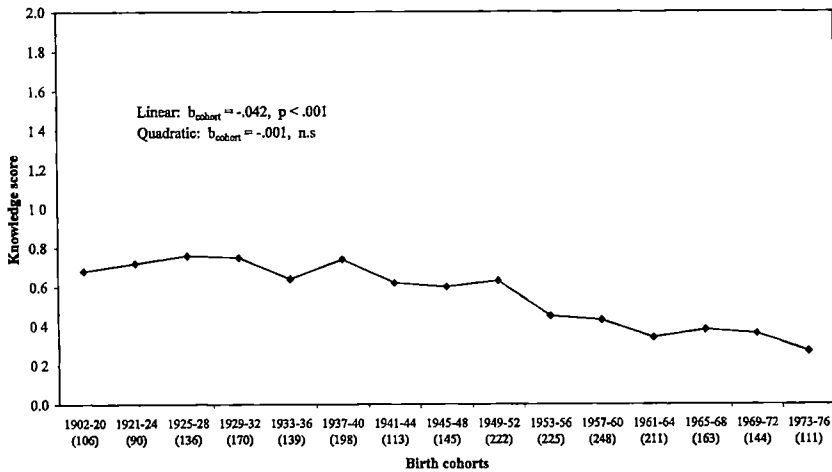


FIG. 10.—Knowledge of the Doctors' Plot (1953) by cohort, controlled for education, gender, and urban/rural location.

other event that has a similarly high level for the oldest cohort is the Prague Spring, which also may have been seen by some as connected to repressive action.) For all other events—Laika, Katya Lycheva, the Cuban Missile Crisis, and to some extent the Virgin Lands Campaign—the oldest cohort is among the less knowledgeable.

The Doctors' Plot

In 1953, Stalin attacked Kremlin physicians with the fabricated claim that they were attempting to destroy party and military leaders. The "Doctors' Plot" was seen at the time as the ominous beginning of a new purge, quite possibly to be as pervasive—had Stalin lived—as the 1936–38 terror (Riasanovsky 1993). The drop in knowledge of the Doctors' Plot shown in figure 10 occurs mainly after the 1949–52 cohort, which is puzzling in terms of period effects since that cohort would have been too young to have experienced even through their parents' eyes what was happening at the national level. Once Stalin was dead, the supposed plot would probably have had fewer long-term reverberations than did Khrushchev's speech to the Twentieth Congress, although we note that overall knowledge levels for these two events are similar. The period effect represented by separation between the two levels of knowledge in figure 10 is highly significant ($P < .001$), and this holds regardless of how the 1949–52 cohort is classified, though we remain unable to explain the level of the 1949–52 cohort as such.

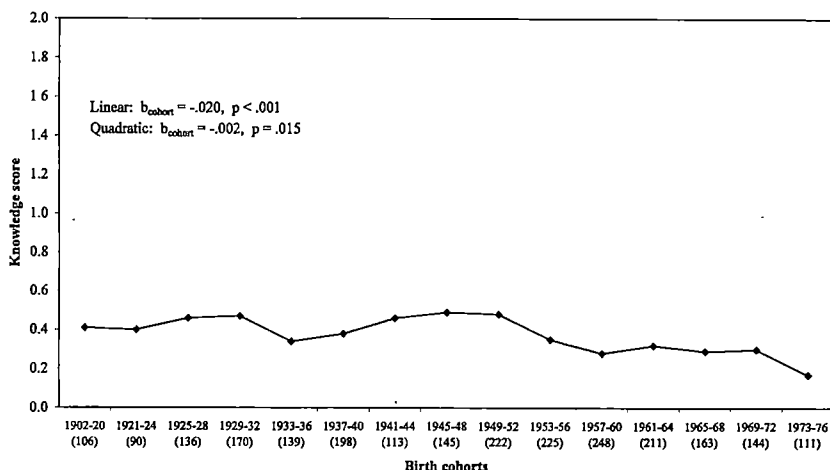


FIG. 11.—Knowledge of the Prague Spring (1968) by cohort, controlled for education, gender, and urban/rural location.

We did not find evidence even for the Yezhovshchina of a change in levels of overall knowledge directly attributable to glasnost, and thus its absence in the case of the Doctors' Plot is not surprising. What we do have for the latter is evidence of the *loss* of important knowledge content by young people. The supposed plot had been characterized by Stalin as a Zionist conspiracy, and it had a clear anti-Semitic emphasis. Mentions of anti-Semitism had been one element in our correctness coding, but only 153 out of the 931 responses coded as correct/partly correct included such a mention. Using these 931 nonzero responses, there is a significant association ($P < .02$) between membership in older cohorts and mentions that Jews were a special target of the Doctors' Plot.³⁵ Evidently, younger Russians simply assimilated the Doctors' Plot to Stalinist purges more generally and were less aware of its anti-Jewish emphasis.

A Final Puzzle: The Prague Spring

The last of our nine events represented the attempted liberalization in 1968 of the oppressive communist regime in Czechoslovakia, which was swiftly put down by Soviet-led military force, justified as essential for the defense of socialism. Alternative views were available during glasnost, but would not have been salient to much of the public at that late point. There appear to be two knowledge peaks for this event in figure 11. The

³⁵ This calculation is based on logistic regression, controlled for education, gender, and urban/rural location. For the analysis we omitted 21 respondents who identified themselves as Jews on a question about nationality.

clearer one is for the 1941 to 1952 cohorts, who were 16–27 at the point that the Soviet Union crushed the attempt by reformers in Czechoslovakia to create “socialism with a human face”; thus this part of the figure can be interpreted as a critical age cohort effect, with the event having maximum impact on those who were young at the time of the Prague Spring. (Using the range between the 1933–36 cohort and the 1957–60 cohort, $P < .001$ for the quadratic term.)

The second slightly lower peak is for the 1925–32 cohorts, which were much older in 1968, well beyond usual definitions of a critical age and an age group not obviously connected to the event in terms of its specific content. If we could regard this peak as due to sampling error, we could treat the higher knowledge of all cohorts born before 1957 as reflecting a more general period effect, as occurs for most other events. However, the quadratic term for the range between 1921 and 1936 reaches significance ($P = .020$), which cautions against dismissing the second peak too easily. Likewise, the drop in knowledge for those in the 1933–40 cohorts cannot readily be treated as chance (the range from 1929–32 to 1941–44 also produces a significant quadratic term). Thus it is necessary to leave this final figure without an adequate and testable interpretation, allowing for the possibility of either an unusual chance occurrence or a substantive explanation not yet discovered.³⁶

The Two Meanings of Education

The strongest associations with knowledge shown earlier in table 2 are those for educational attainment, but we proposed that this variable be conceptualized as having two different meanings.³⁷ The high associations, we believe, reflect not the content of schooling, but rather the greater knowledge of the world open to those with intellectual and informational resources and interests that enabled them to learn beyond what was officially available to the general Russian public.

³⁶ There is one other finding with regard to both the Prague Spring and the Cuban Missile Crisis, the two items that focused on events outside the Soviet Union. For most events we find a regular ordering of knowledge with Moscow at the top, St. Petersburg second, other urban areas third, and rural areas least knowledgeable. But for these two foreign events, St. Petersburg residents show as much (the Prague Spring) or more (Cuban Missile Crisis) knowledge than do Muscovites. Such greater knowledge of foreign happenings is consistent with St. Petersburg's historical westward orientation, as well as with its tendency to support radical reformers (e.g., in the 1989 election).

³⁷ Note that table 2 presents estimated zero-order and adjusted magnitudes of association but does not show direction or assume linearity; however, graphing indicates that all the relations involving education are clearly positive and reveal only trivial departures from monotonicity.

Thus the strongest association (both η and β in table 2) is with knowledge of Solzhenitsyn's harshly realistic novel *One Day in the Life of Ivan Denisovich*, the least-known event in table 1. This is not simply a matter of specialized interest in serious literature, for close behind it in size of associations with education is knowledge of the Doctors' Plot.³⁸ Neither of these events was treated in schools or in the mass media in the period after its initial occurrence and prior to the advent of glasnost.

In its other meaning, education refers to the deliberate socialization of the young. Where primary and secondary levels of schooling are widespread and centrally controlled, as was true in the Soviet Union, such socialization, importantly aided by the controlled mass media, should show up in the form of low levels of association between years of education and knowledge of events deemed favorable to the country's image. Virtually the entire population will have been encouraged to learn about such events. This applies to three events in table 1 that were enthusiastically promoted by the regime: Laika, Katya Lycheva, and the Virgin Lands Campaign. These three present the lowest associations with years of schooling (both η and β in table 2), but this indicates the pervasiveness of direct socializing effects rather than their absence.³⁹ More generally, the greater the knowledge shown for an event in table 1, the less is its association with years of schooling: the correlation (r) between the sizes of the weighted knowledge scores in table 1 and the sizes of the nine β s for education in table 2 is $-.61$ ($P = .079$; $N = 9$), with the correlation using nonzero scores almost the same ($-.59$).

We cannot disentangle media and school effects very well, but since some effects on knowledge are concentrated in a very young part of the population (Laika) and other effects show relatively little specificity by age (the Virgin Lands Campaign), we suspect that teaching through the media is at least as important as teaching in schools.

CONCLUSION

Our study focused on two factors that influenced knowledge of past events by ordinary Russians in 1994: their education, represented by years of

³⁸ A significant cohort-education interaction ($P < .001$) when graphed indicates that it is lower educational levels whose knowledge appears especially traceable to the time of Stalin's announcement of the Doctors' Plot; at higher educational levels, knowledge is not so tied to the date of Stalin's fabricated claim.

³⁹ This cannot be the whole story, since especially for Laika and Katya Lycheva there are many Russians in the critical cohorts who were scored zero. However, urban/rural location and advanced education also contribute to scores; for Katya Lycheva, e.g., ignorance disappears almost completely among the small number of young, highly educated Muscovites.

schooling, and their location in time, represented by birth cohort. Neither is as simple as it appears on first consideration.

Education

Events like the dog Laika's orbit in *Sputnik 2* were favorable to the reputation of the Soviet regime and therefore publicized by the state-controlled media and schools. Knowledge of these events has a relatively low correlation with years of schooling, because most Russians were exposed to them early on regardless of their eventual educational attainment. Other events, like Solzhenitsyn's novel, suppressed over a long period by the regime, show stronger associations with educational attainment. This is not because they were studied at higher levels of schooling, but because individuals who achieved advanced education could develop the critical capacity and resources to gain information about the world in ways not sanctioned by the regime. Individuals with high levels of education are less limited to their own personal experience, and in this sense education acts to reduce cohort effects.

Our distinction between two meanings of education is a relative one, for all events show a significant and monotonic association between years of schooling and knowledge. But the distinction suggests that even in a society where the state controls virtually all forms of media and instruction, those who develop greater cognitive sophistication can acquire knowledge beyond and around officially available information. This should be even more true today with the increasing difficulty of closing a country's borders to satellite broadcasting and to the Internet, both of which are apt to be most accessible to those with greater education. Moreover, to the extent that greater cognitive sophistication develops, restraints are placed on the kind of elite manipulation of collective memory that is described by historians like Bodnar (1992) and Hobsbawm and Ranger (1983).

Age

Our findings about age-related effects are more complex and require consideration of three distinct processes, though more than one might operate at the same time. One process is essentially a period effect: individuals alive when an event occurs are more knowledgeable than those born even soon after the event has ended. In our data, we find evidence that this process applies to knowledge of Laika, the Cuban Missile Crisis, *One Day in the Life of Ivan Denisovich*, the Yezhovshchina, the Twentieth Congress of the CPSU, and the Doctors' Plot. Its application to the Prague

Spring is more uncertain, and it is difficult to apply at all to Katya Lycheva and the Virgin Lands Campaign, since no respondents in our sample were born after the two events ended.

Cohort effects of two different kinds point to the other two processes. One, which is limited but clear in our data, is the intrinsic interest that certain events hold for particular age groups. This is shown most dramatically for two events, both of which were simple in nature and highly publicized by the government: the dog Laika, known especially by those who were very young children at the time, and the girl diplomat Katya Lycheva, where knowledge is highest among those who were adolescent girls at the time. Earlier findings on Americans' knowledge of the Vietnam War and of Woodstock also point to the importance of event content for particular ages (Jennings 1996; Schuman et al. 1997).

Our strongest evidence for the other process—adolescence and early adulthood as a generic critical developmental stage for acquiring knowledge and therefore producing later cohort effects—occurs with the novel *One Day in the Life of Ivan Denisovich*. The publication of Solzhenitsyn's writings toward the end of glasnost seems to have aroused interest and created knowledge specifically for young Russians. Since neither the novel nor other works by Solzhenitsyn had special relevance to youth as such, it seems likely that knowledge in this case reflects the classic primacy process discussed by Mannheim. We can state as a further proposition that it is primarily at a point of radical change that new ideas are especially appealing to youth. During such periods, older people can still cling for a time to the world as they knew it, but those too young to be anchored in the past are open to learning quickly from new events. As Swidler (1986) argues, it is precisely the young who bear the lowest costs of acquiring new "strategies of action" that may be symbolized by cultural knowledge during unsettled times.⁴⁰

We also found evidence for more subtle cohort effects in the detailed content of knowledge; for example, even though the Great Purge of the 1930s is widely known among Russians from different cohorts, the older Russians we interviewed focused more on those who were victims of the Purge and younger Russians more on those who were the perpetrators. Subtle cohort differences within larger period effects occurred as well for knowledge of the Virgin Lands Campaign and of the Doctors' Plot.

⁴⁰ A further example by Norbert Elias that is analogous to the Russian case concerns the shift toward "informalization" by young Germans to distinguish themselves from Germans who had lived through World War II and the end of the Third Reich (Goudsblom and Mennell 1998).

Larger Theoretical Issues

Our findings have implications for several larger issues. First, we have treated our study of collective knowledge as part of broader research efforts on collective memory, yet our results show less evidence of the importance of a critical age than was true of some previous research (e.g., Schuman and Scott 1989; Scott and Zac 1993). There is a significant difference, however, between the two types of inquiry. The earlier work on collective memories allowed respondents to produce their own spontaneous memories of any public events that they personally considered important. Our research on knowledge called for correct accounts of events that *we* chose and presented, regardless of whether they would have occurred to respondents and irrespective of when they happened in the course of or even before a respondent's own lifetime. Thus, personal memories may refer back primarily to experiences during one's adolescence or early adulthood, while objective knowledge more easily spans all the events experienced over a lifetime, as well as reaching backward to a past not directly experienced by the individual.

This distinction suggests that although the term "collective memory" is useful for conceptualizing the broad area into which our and other research falls, the term does not point unambiguously to particular phenomena at either the individual or the cultural level. At the individual level, what are produced as memories, shared or otherwise, are always in part a function of the type of questions posed and of the setting in which the remembering takes place (e.g., as Halbwachs would urge, who else is present). At the cultural level, it is at least equally difficult to specify a unique collective memory: what are said to be "publicly available symbols" are inevitably a function of what particular researchers define as "publicly available." Thus, some information about the Doctors' Plot was available to some Russians at the time it occurred, as well as later during Khrushchev's Thaw, and it was also brought up again openly during glasnost; yet it was not effectively "available" to those too preoccupied with economic survival to be attentive, or to those living in rural areas cut off from discussions in urban Russia, or to others for many different reasons. Both at the individual and the cultural level, what is seen to be a particular collective memory is always partly—though certainly not wholly—a function of decisions made by the researcher.

Certain of our results also have implications for another issue involving collective memory: the debate over "presentism"—the question of how much our knowledge about the past is constructed to meet present needs and problems, without much concern for objective information about past events (Schwartz 1990). We found that at least for older Russians, the past was retained in their memories of Khrushchev's role in the Virgin

Lands Campaign, even though during the long years after his ouster the Soviet government made an effort to connect the campaign exclusively to Brezhnev. Such official efforts at redefinition had their major impact on young people learning about the campaign for the first time, but did not change the knowledge of those whose learning dated from an earlier point. Similarly, older Russians retained knowledge of the anti-Semitic emphasis of the Doctors' Plot, but younger Russians assimilated the plot to Stalinist purges more generally and were ignorant of its anti-Semitic focus. Thus, crucial information about the past is most easily lost when cohorts who directly experienced an event are replaced by new cohorts who can learn about it only secondhand.

Does our study of the knowledge of a cross-section of Russians tell us anything about their attitudes and actions toward a new event? Sociologists working in the area of neoinstitutionalism call for connecting norms and beliefs to actions in institutional contexts (Nee 1998). We were able to investigate one such possible connection: that between knowledge of the Great Purge of the 1930s and self-reported voting preference in the 1993 Russian parliamentary election. Virtually all correct/partly correct responses about the Great Purge had treated it as a time of persecution and repression; hence we hypothesized that those who showed the most knowledge of that traumatic time should have been most likely to favor democratically oriented parties in the 1993 election. We used logistic regression to predict voting preference for the most democratic parties ($N = 410$) versus all other positions ($N = 2,011$ for centrists, communists, nationalists, and nonvoters combined), controlling for age, education, gender, and urban/rural location, and found the hypothesized relation ($P < .03$). If the democratic and centrist parties are combined in order to increase the N of voting in a relatively democratic direction, the relation is still more reliable ($P < .001$). Although any such correlational result is open to more than one causal interpretation, the finding is consistent with the proposition that knowledge of this single but important event from the 1930s had implications for behavior in an institutional context in 1993.

Finally, we should acknowledge that our conclusions about collective knowledge are based on asking about nine specific events in one country at one point in time, albeit with some support from previous studies. How far the results can be generalized to other cases remains to be seen. But in emphasizing the importance of personal experience in all three types of age-related processes, we return to the spirit of Mannheim's larger theoretical exposition of generational effects. More crucial to his approach than the focus on a "critical age" was the distinction of "generation as an actuality" from generation as "a mere location phenomenon" (1952, p. 303)—with the former requiring "participation in [a] common destiny." That distinction invites the emphasis on event meaning that we have used

to modify a purely developmental approach to understanding how knowledge is acquired and remembered.

REFERENCES

- Andrews, Frank M., James N. Morgan, John A. Sonquist, and Laura Klem. 1973. *Multiple Classification Analysis*. Ann Arbor, Mich.: Institute for Social Research.
- Barthel, Diane. 1996. *Historic Preservation: Collective Memory and Historical Identity*. New Brunswick, N.J.: Rutgers University Press.
- Bernstein, Michael André. 1994. *Foregone Conclusions: Against Apocalyptic History*. Berkeley and Los Angeles: University of California Press.
- Bodnar, John. 1992. *Rethinking America: Public Memory, Commemoration, and Patriotism in the Twentieth Century*. Princeton, N.J.: Princeton University Press.
- Brinton, Mary C., and Victor Nee, eds. 1998. *The New Institutionalism in Sociology*. New York: Russell Sage Foundation.
- Brown, Roger, and James Kulik. 1977. "Flashbulb Memories." *Cognition* 5:73-99.
- Converse, Philip E. 1987. "The Enduring Impact of the Vietnam War on American Public Opinion." Pp. 53-75 in *After the Storm: American Society a Decade after the Vietnam War*. Taipei: Institute of American Culture.
- Delli Carpini, Michael X., and Scott Keeter. 1996. *What Americans Know about Politics and Why It Matters*. New Haven, Conn.: Yale University Press.
- Eliot, T. S. 1943. *Four Quartets*. New York: Harcourt, Brace.
- Erikson, Erik. 1950. *Childhood and Society*. New York: W. W. Norton.
- Fentress, James, and Chris Wickham. 1992. *Social Memory*. Oxford: Blackwell.
- Fields, James M., and Howard Schuman. 1976. "Public Beliefs about the Beliefs of the Public." *Public Opinion Quarterly* 40:427-48.
- Firebaugh, Glenn, and Kevin Chen. 1995. "Vote Turnout of Nineteenth Amendment Women: The Enduring Effect of Disenfranchisement." *American Journal of Sociology* 100:972-96.
- Fitzgerald, J. M., and R. Lawrence. 1984. "Autobiographical Memory across the Life-Span." *Journal of Gerontology* 39:692-99.
- Gardner, Eileen M. 1988. "Teaching History in the U.S. and USSR." *Education Digest* 54:29-32.
- Garthoff, Raymond. 1989. *Reflections on the Cuban Missile Crisis*. Washington, D.C.: Brookings Institution.
- Gillis, John. 1994. "Memory and Identity: The History of a Relationship." Pp. 3-24 in *Commemorations: The Politics of National Identity*. Princeton, N.J.: Princeton University Press.
- Gorbachev, Mikhail Sergeevich. 1996. *Memoirs*. New York: Doubleday.
- Goudsblom, Johan, and Stephen Mennell. 1998. *The Norbert Elias Reader*. Oxford: Blackwell.
- Greenstein, Fred. 1965. *Children and Politics*. New Haven, Conn.: Yale University Press.
- Greif, Avner. 1998. "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies." Pp. 77-104 in *The New Institutionalism in Sociology*, edited by Mary C. Brinton and Victor Nee. New York: Russell Sage Foundation.
- Halbwachs, Maurice. (1925) 1992. "The Social Framework of Memory." Pp. 37-189 in *On Collective Memory*, edited by Lewis A. Coser. Chicago: University of Chicago Press.
- . (1950) 1980. *The Collective Memory*. New York: Harper.
- Hobsbawm, Eric. 1987. *The Age of Empire: 1875-1914*. New York: Pantheon Books.

- Hobsbawm, Eric, and Terence Ranger. 1983. *The Invention of Tradition*. Cambridge: Cambridge University Press.
- Holmes, Alison, and Martin A. Conway. 1999. "Generational Identity and the Reminiscence Bump: Memories for Public and Private Events." *Journal of Adult Development* 6:21-34.
- Inkeles, Alex, and Raymond A. Bauer. 1959. *The Soviet Citizen: Daily Life in a Totalitarian Society*. Cambridge, Mass.: Harvard University Press.
- Irwin-Zarecka, Iwona. 1994. *Frames of Remembrance: The Dynamics of Collective Memory*. New Brunswick, N.J.: Transaction Publishers.
- Jaeger, Hans. (1977) 1985. "Generations in History: Reflections on a Controversial Concept." *History and Theory* 24:273-92.
- Jennings, M. Kent. 1996. "Political Knowledge across Time and Generations." *Public Opinion Quarterly* 60:228-52.
- Lang, Kurt, and Gladys Engel Lang. 1978. "Experiences and Ideology: The Influence of the Sixties on an Intellectual Elite." Pp. 197-230 in *Social Movements: Conflicts and Change*, vol. 1. Edited by L. Kriesberg. Greenwich, Conn.: JAI Press.
- . 1983. *The Battle for Public Opinion: The President, the Press, and the Polls during Watergate*. New York: Columbia University Press.
- . 1990. "Collective Memory and the News." Pp. 19-35 in *Mass Communication and Political Information Processing*, edited by Sidney Kraus. Hillsdale, N.J.: Erlbaum.
- Lang, Kurt, Gladys Engel Lang, Hans-Mathias Kepplinger, and Simone Ehmgig. 1993. "Collective Memory and Political Generations: A Survey of German Journalists." *Political Communication* 10:211-29.
- Larson, S. F. 1996. "Memorable Books: Recall of Reading and Its Personal Context." Pp. 583-99 in *Empirical Approaches to Literature and Aesthetics*, edited by Roger J. Kreuz and Mary Sue MacNealy. Norwood, N.J.: Ablex.
- Mackavay, W. R., J. E. Malley, and A. J. Stewart. 1991. "Remembering Autobiographically Consequential Experiences: Content Analysis of Psychologists' Accounts of Their Lives." *Psychology and Aging* 6:50-59.
- Mannheim, Karl. (1928) 1952. "The Problem of Generations." Pp. 276-322 in *Essays on the Sociology of Knowledge*, by Karl Mannheim. London: Routledge & Kegan Paul.
- Nee, Victor. 1998. "Sources of the New Institutionalism." Pp. 1-16 in *The New Institutionalism in Sociology*, edited by Mary C. Brinton and Victor Nee. New York: Russell Sage Foundation.
- Nora, Pierre. 1996. "Generation." Pp. 499-531 in *Realms of Memory: Rethinking the French Past*, vol. 1: *Conflicts and Divisions*. Edited by Lawrence D. Kritzman. New York: Columbia University Press.
- . 1996-98. *Realms of Memory: Rethinking the French Past*, 3 vols. Edited by Lawrence D. Kritzman. New York: Columbia University Press.
- Olick, Jeffrey K., and Joyce Robbins. 1998. "Social Memory Studies: From 'Collective Memory' to the Historical Sociology of Mnemonic Practices." *Annual Review of Sociology* 25:105-40.
- Pennebaker, James W., and Becky L. Banasik. 1997. "On the Creation and Maintenance of Collective Memories." Pp. 3-19 in *Collective Memory of Political Events: Social Psychological Perspectives*, edited by J. W. Pennebaker, D. Paez, and B. Rimé. Mahwah, N.J.: Erlbaum.
- Peterson, Merrill D. 1977. *The Portable Thomas Jefferson*. New York: Penguin Books.
- Pillemer, David B. 1998. *Momentous Events, Vivid Memories*. Cambridge, Mass.: Harvard University Press.
- Remnick, David. 1993. *Lenin's Tomb: The Last Days of the Soviet Empire*. New York: Random House.
- Riasanovsky, Nicholas V. 1993. *A History of Russia*. New York: Oxford University Press.

- Rieder, Jonathan. 1994. "Doing Political Culture: Interpretive Practice and the Earnest Heuristic." Pp. 117–51 in *Political Culture and Political Structure: Theoretical and Empirical Studies*, edited by Frederick D. Weil and Mary Gautier. Vol. 2 of *Research on Democracy and Society*. Greenwich, Conn.: JAI Press.
- Rieger, Cheryl Gene. 1995. "Collective Memories in the Late Soviet Union." Ph.D. dissertation. University of Michigan, Department of Sociology.
- Roberts, Carl W., and Kurt Lang. 1985. "Generations and Ideological Change: Some Observations." *Public Opinion Quarterly* 49:460–73.
- ROMIR. 1994. *Technical Report*. Russian Public Opinion and Market Research (ROMIR).
- Rubin, David C., Scott E. Wetzler, and Robert D. Nebes. 1986. "Autobiographical Memory across the Lifespan." Pp. 202–21 in *Autobiographical Memory*, edited by David C. Rubin. Cambridge: Cambridge University Press.
- Rubin, David D., Tamara A. Rahhal, and Leonard W. Poon. 1998. "Things Learned in Early Adulthood Are Remembered Best." *Memory and Cognition* 26:3–19.
- Ryder, Norman B. 1965. "The Cohort in the Study of Social Change." *American Sociological Review* 30:843–61.
- Schudson, Michael. 1989. "How Culture Works." *Theory and Society* 18:153–89.
- . 1992. *Watergate in American Memory*. New York: Basic Books.
- Schuman, Howard, Hiroko Akiyama, and Bärbel Knäuper. 1998. "Collective Memories of Germans and Japanese about the First Half Century." *Memory* 6:427–54.
- Schuman, Howard, Robert F. Belli, and Katherine Bischooping. 1997. "The Generational Basis of Historical Knowledge." Pp. 47–77 in *Collective Memories of Political Events: Social Psychological Perspectives*, edited by J. W. Pennebaker, D. Paez, B. Rimé. Mahwah, N.J.: Erlbaum.
- Schuman, Howard, and Jacqueline Scott. 1987. "Problems in the Use of Survey Questions to Measure Public Opinion." *Science* 236: 957–59.
- . 1989. "Generations and Collective Memories." *American Sociological Review* 54:359–81.
- Schwartz, Barry. 1982. "The Social Context of Commemoration: A Study in Collective Memory." *Social Forces* 61:374–402.
- . 1990. "The Reconstruction of Abraham Lincoln." Pp. 81–107 in *Collective Remembering*, edited by David Middleton and Derek Edwards. London: Sage.
- . 1996. "Memory as a Cultural System: Abraham Lincoln in World War II." *American Sociological Review* 61:908–27.
- . 1998. "Frame Images: Towards a Semiotics of Collective Memory." *Semiotica* 121½:1–40.
- . 1999. "Memory and the Practices of Commitment." Pp. 135–46 in *Sociology as Everyday Life*, edited by Barry Glassner and Rosanna Hertz. Beverly Hills: Sage.
- Scott, Jacqueline, and Lilian Zac. 1993. "Collective Memories in Britain and the United States." *Public Opinion Quarterly* 57:315–31.
- Sears, David O., and Nicholas A. Valentino. 1997. "Politics Matters: Political Events as Catalysts for Preadult Socialization." *American Political Science Review* 91:45–65.
- Sigel, Roberta S. 1965. "An Exploration into Some Aspects of Political Socialization of School Children's Reactions to the Death of a President." Pp. 34–69 in *Children and the Death of a President*, edited by Martha Wolfenstein and Gilbert Kliman. Garden City, N.Y.: Anchor Books.
- Skinner, C. J., D. Holt, and T. M. F. Smith. 1989. *Analysis of Complex Surveys*. New York: Wiley.
- Sturkin, Marita. 1997. *Tangled Memories: The Vietnam War, the AIDS Epidemic, and the Politics of Remembering*. Berkeley and Los Angeles: University of California Press.
- Swidler, Ann. 1986. "Culture in Action." *American Sociological Review* 51:273–86.

- Swidler, Ann, and Jorge Ardití. 1994. "The New Sociology of Knowledge." *Annual Review of Sociology* 20:305–29.
- Tolz, Vera. 1990. *The USSR in 1989: A Record of Events*. Boulder, Colo.: Westview Press.
- Whitney, Thomas, ed. 1963. *Khrushchev Speaks: Selected Speeches, Articles and Press Conferences, 1949–1961*. Ann Arbor: University of Michigan Press.
- Zelizer, Barbie. 1995. "Reading the Past against the Grain: The Shape of Memory Studies." *Critical Studies in Mass Communications* 12:214–39.
- Zerubavel, Eviatar. 1997. *Social Mindscales: An Invitation to Cognitive Sociology*. Cambridge, Mass.: Harvard University Press.
- Zimmerman, William. 1987. "Mobilized Participation and the Nature of the Soviet Dictatorship." Pp. 332–53 in *Politics, Work, and Daily Life in the Soviet Dictatorship: A Survey of Former Soviet Citizens*, edited by James R. Millar. Cambridge: Cambridge University Press.

Working-Class Power, Capitalist-Class Interests, and Class Compromise¹

Erik Olin Wright
University of Wisconsin

This article proposes a general theoretical framework for understanding the concept of "class compromise" in terms of a "reverse-J" model of the relationship between the *associational power* of workers and the *interests* of capitalists: increases in working-class power adversely affect capitalist-class interests until such power crosses some intermediate threshold beyond which further increases in working-class power are potentially beneficial to capitalists' interests. This article argues that the reverse-J curve is itself the result of two distinct kinds of effects of workers' power on capitalists' interests: one, a negative effect, in which workers' power undermines the capacity of capitalists to unilaterally make various kinds of decisions, and the second, a positive effect, in which workers' power helps capitalists solve the various kinds of collective action problems they face.

The concept of "class compromise" invokes three quite distinct images. In the first, class compromise is an illusion. Leaders of working-class organizations—especially unions and parties—strike opportunistic deals with the capitalist class that promise general benefits for workers but that, in the end, are largely empty. Class compromises are, at their core, one-sided capitulations rather than reciprocal bargains embodying mutual concessions.

In the second image, class compromises are like stalemates on a battlefield. Two armies of roughly similar strength are locked in battle. Each is sufficiently strong to impose severe costs on the other; neither is strong enough to definitively vanquish the opponent. In such a situation of stalemate, the contending forces may agree to a "compromise": to refrain from mutual damage in exchange for concessions on both sides. The concessions

¹ I would like to thank Samuel Bowles, Andrew Mitchner, Michael Burawoy, Joel Rogers, and Robert Boyer for their extremely constructive criticisms of earlier drafts of this article. Direct correspondence to Erik Olin Wright, Department of Sociology, University of Wisconsin, 1180 Observatory Drive, Madison, Wisconsin 53706. E-mail: Wright@ssc.wisc.edu

are real, not phoney, even if they are asymmetrical. Still, they do not constitute a process of real cooperation between opposing class forces. This outcome can be referred to as a "negative class compromise."

The third image sees class compromise as a form of mutual cooperation between opposing classes. This is not simply a situation of a balance of power in which the outcome of conflict falls somewhere between a complete victory or a complete defeat for either party. Rather, here there is a possibility of a non-zero-sum game between workers and capitalists, a game in which both parties can improve their position through various forms of active, mutual cooperation. This outcome can be called a "positive class compromise."

The basic objective of this article is to explore the theoretical logic of positive class compromises and to propose a general model of the conditions conducive to them in developed capitalist societies. The article will not attempt a systematic empirical investigation, although empirical illustrations will be used to clarify elements of the model. The premise of the analysis is that so long as capitalism in one form or another is the only historically available way of organizing an economy, a positive class compromise—if it is achievable—will generally constitute the most advantageous context for the improvement of the material interests and life circumstances of ordinary people. If one is interested in advancing such interests, therefore, it is important to understand the conditions that facilitate or hinder the prospects for positive class compromise.

The central argument I will make is that the possibilities for stable, positive class compromise generally hinge on the relationship between the *associational power* of the working class and the *material interests* of capitalists. The conventional wisdom among both neoclassical economists and traditional Marxists is that, in general, there is an inverse relationship between these two variables: increases in the power of workers adversely affect the interests of capitalists (see fig. 1). The rationale for this view is straightforward for Marxist scholars: since the profits of capitalists are closely tied to the exploitation of workers, the material interests of workers and capitalists are inherently antagonistic. Anything that strengthens the capacity of workers to struggle for and realize their interests, therefore, negatively affects the interests of capitalists. The conventional argument by neoclassical economists is somewhat less straightforward, for they deny that in a competitive equilibrium workers are exploited by capitalists. Nevertheless, working-class associational power is seen as interfering with the efficient operation of labor markets by making wages harder to adjust downward when needed and by making it harder for employers to fire workers. Unions and other forms of working-class power are seen as forms of monopolistic power within markets, and like all such practices, they generate monopoly rents and inefficient allocations. As a result, unionized

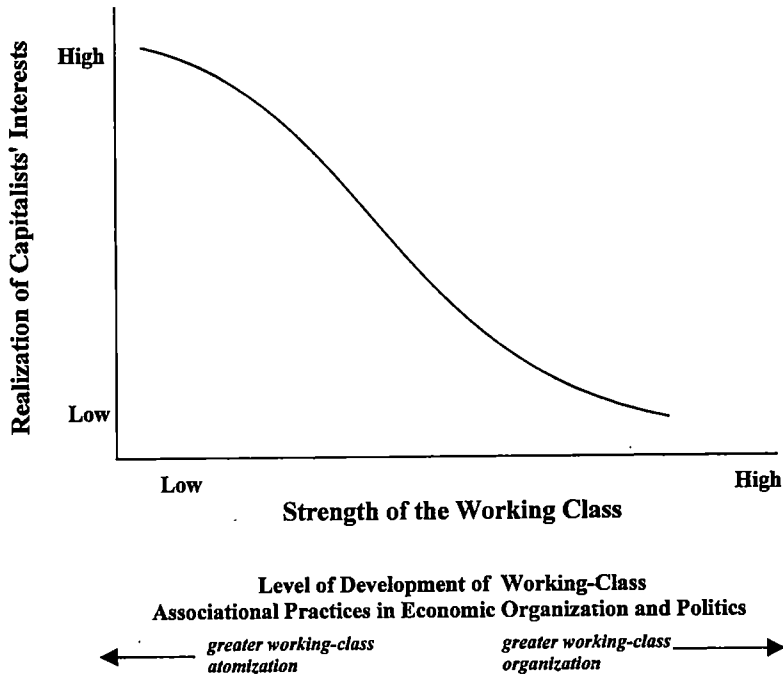


FIG. 1.—Conventional view of the relationship between working-class strength and capitalist-class interests.

workers are able to extort a monopoly rent in the form of higher wages at the expense of both capitalists and nonunionized workers.

This article explores an alternative understanding of the relationship between workers' power and capitalists' interests: instead of an inverse relationship, this alternative postulates a curvilinear *reverse-J* relationship (see fig. 2).² As in the conventional wisdom, capitalist-class interests are best satisfied when the working class is highly disorganized: when workers compete with each other in an atomized way and lack significant forms of associational power. As working-class power increases, capitalist-class interests are initially adversely affected. However, once working-class power crosses some threshold, working-class associational power begins to have positive effects on capitalists' interests. As we shall see in more detail below, these conditions allow for significant gains in productivity and rates of profit due to such things as high levels of bargained

² The reverse-J shaped relationship between working-class power and capitalists' interests was first suggested to me in an article by Joel Rogers (1990).

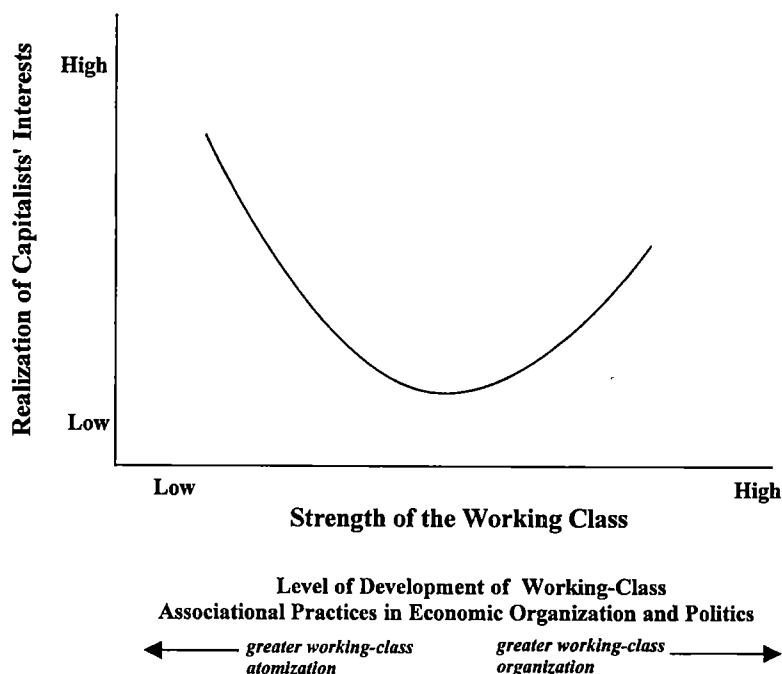


FIG. 2.—Curvilinear relationship between working-class strength and capitalist-class interests.

cooperation between workers and capitalists, rationalized systems of skill upgrading and job training, enhanced capacity for solving macroeconomic problems, and a greater willingness of workers to accept technological change given the relative job security they achieve because of union protections. The upward-bending part of the curve, where increases in working-class power have positive effects on capitalist-class interests, generates conditions for positive class compromise. The goal of this article, then, is to elaborate a general theoretical model of the causal processes underlying the relation presented in figure 2.

The first section of this article briefly defines the core concepts used in the analysis. The second section situates the problem of positive class compromise within a broader literature on interclass cooperation, labor relations, and economic governance. The next section then frames the problem of class compromise in terms of various possible game-theory models of the interactions of workers and capitalists. With this game-theoretic background, the fourth section elaborates general theoretical arguments about the underlying mechanisms for the reverse-J model of positive class compromise. The last section concludes the article with a

somewhat speculative discussion of the impact of globalization on the prospects of class compromise.

CORE CONCEPTS AND METHODOLOGICAL ISSUES

Some of the concepts used in this article do not have transparent meanings. In particular, the concepts of "class," "interests," and "power" are all highly contested. I will not attempt to elaborate analytically precise definitions of any of these concepts here, but some brief clarifications are necessary.

Class.—The meaning of class and its related concepts—class structure, class struggle, class formation, class compromise—can be analyzed at various levels of abstraction. For some purposes, it is important to deploy a highly differentiated class concept that elaborates a complex set of concrete locations within class structures. My work on the problem of the "middle class" and "contradictory locations within class relations" would be an example of such an analysis (Wright 1985, 1997). For some problems, the causal processes cannot be properly studied without specifying a range of fine-grained differentiations and divisions within classes on the basis of such things as sector, status, gender, and race. For other purposes, however, it is appropriate to use a much more abstract, simplified class concept, revolving around the central polarized class relation of capitalism: capitalists and workers. This is the class concept I will use in most of this article.

In a stylized Marxian manner, I will define capitalists as those people who own and control the capital used in production and workers as all employees excluded from such ownership and control. In this abstract analysis of class structure, I will assume that these are mutually exclusive categories. There is thus no middle class as such. No workers own any stock. Executives, managers, and professionals in firms are either amalgamated into the capitalist class by virtue of their ownership of stock and command of production, or they are simply part of the "working class" as employees. Of course, this is unrealistic. My claim, however, is that this abstract, polarized description of class relations in capitalism can still be useful in clarifying real mechanisms that actual actors face and is thus a useful point of departure for developing a theory of class compromise.³

³ The claim that this abstract polarized concept of class is analytically useful may be controversial. Given that actual capitalist firms engaging in complex strategies and bargaining with their employees encounter considerable heterogeneity, and this heterogeneity in fact does matter for the optimal profit maximizing strategy of the firm, it may seem illegitimate to bracket such complexity in favor of a simple model of capital and labor. As in all attempts at elaborating theoretical models, the appropriate level of abstraction is a matter of contestation.

Interests.—Throughout this article, our attention will be restricted to what can be narrowly termed the “material interests” of people by virtue of their class location, or what I will refer to in shorthand as “class interests.”⁴ In general, I will make two radical simplifying assumptions about the nature of these interests: first, that class interests can be reduced to a single quantitative dimension so that one can talk about the extent to which the interests of the members of a class are realized; and second, that all people in a given class location share the same class interests. Both of these assumptions are problematic when we study concrete capitalist societies but, as in the adoption of a simple polarized class structure concept, are useful simplifications for the present analytical purposes.

Power.—Like “interests,” “power” is used in many different ways in social theory. In the context of class analysis, power can be thought of as the *capacity of individuals and organizations to realize class interests*. Insofar as the interests of people in different classes—say workers and capitalists—are opposed to each other, this implies that the capacity of workers to realize their class interests depends in part on their capacity to counter the power of capitalists. Power, in this context, is thus a relational concept.

In this article, our concern is mainly with what I will term working-class “associational” power—the various forms of power that result from the formation of collective organizations of workers. This includes such things as unions and parties but may also include a variety of other forms, such as works councils or forms of institutional representation of workers on boards of directors in schemes of worker codetermination, or even, in certain circumstances, community organizations. Associational power is to be contrasted with what can be termed “structural power”—power that results simply from the location of workers within the economic system. The power of workers as individuals that results directly from tight labor markets or from the strategic location of a particular group of workers within a key industrial sector would constitute instances of structural power. While such structural power may itself influence associational power, I will focus in this article on associational power as such.

The models examined here do not directly concern the role of associational power of capitalists in the formation of class compromise. As the literature on neocorporatism has pointed out, there are certain institu-

⁴ When I speak of “class interests,” I will always mean the interests of people determined by their location in the class structure. I do not believe that classes as collective entities have “interests” in a literal sense. Of course, individuals may have interests in the strength of the collective organizations of classes, and the class interests of individuals may be contingent upon the security of the interests of other members of the same class, but this still does not mean that classes as collectivities have “interests.”

tional settings of class compromise in which the associational power of employers plays a pivotal role (Streeck and Schmitter 1985; Streeck 1992; Pontusson 1997), and there will be places in the discussion where reference will be made to the role of such associations. Our concern, however, is not with capitalists' associational power, as such, but with the ways in which working-class associational power has an impact on the interests of capitalists.

There is no implication in the analysis that workers' associational power is entirely exogenous to the processes investigated. While I will concentrate on the ways in which increases in workers' power can positively benefit capitalists' interests, it may also be the case that part of the explanation for the level of workers' associational power is precisely the existence of this beneficial effect and, conversely, that the erosion of such beneficial effects may itself contribute to the decline of workers' power. To the extent that the intensity of resistance by capitalists, individually and collectively, to workers' attempts at creating and sustaining associational power itself partially depends upon the potential benefits to capital of such power, the extent of workers' power will in part be a function of capitalists' interests. In any case, this article does not attempt to develop a dynamic theory of the causes of working-class power but merely of the effects of such power on capitalists' interests.

Sites of Class Compromise

Class struggle and compromise do not occur within an amorphous "society" but within specific institutional contexts—firms, markets, states. The real mechanisms that generate the reverse-J curve in figure 2 are embedded in such institutional contexts. Three institutional spheres within which class struggles occur and class compromises are forged are particularly important and are defined below.

The sphere of exchange.—This concerns, above all, the labor market and various other kinds of commodity markets. In some situations, financial markets may also be an arena within which class conflicts occur and class compromises are forged.

The sphere of production.—This concerns what goes on inside of firms once workers are hired and capital invested. Conflicts over the labor process and technology are the characteristic examples.

The sphere of politics.—Class conflict and class compromise also occur within the state over the formation and implementation of state policies and the administration of various kinds of state-enforced rules.

There is a rough correspondence between each of these institutional spheres of class conflict and class compromise and characteristic kinds of working-class collective organizations: *labor unions* are the characteristic

associational form for conflict and compromise in the sphere of exchange; *works councils* and related associations are the characteristic form within the sphere of production; and *political parties* are the characteristic form within the sphere of politics.

The central task of our analysis, then, is to examine the mechanisms that enable these different forms of working-class associational power—unions, works councils, parties—to forge *positive* class compromises within the spheres of exchange, production, and politics.

SITUATING THE CONCEPT OF CLASS COMPROMISE

In the most abstract and general terms, class compromise—whether positive or negative—can be defined as a situation in which some kind of *quid pro quo* is established between conflicting classes in which, in one way or another, people in each class make “concessions” in favor of the interests of people in the opposing class.⁵ The “compromise” in class compromise is a compromise of class-based interests—members of each class give up something of value. Class compromise is thus always defined against a counterfactual in which such concessions are not made. Typically, this is a situation in which the use of threats, force, and resistance plays a more prominent active role in class interactions.

Defined in this way, the idea of class compromise is closely linked to Gramsci's (1971) concept of “hegemony.” Gramsci uses the concept of hegemony to distinguish two general conditions of capitalist society. In a *nonhegemonic* system, capitalist-class relations are reproduced primarily through the direct, despotic use of coercion. In a hegemonic system, in contrast, class relations are sustained in significant ways through the *active consent* of people in the subordinate classes. Coercion is still present as a background condition—hegemony is “protected by the armor of coercion” in Gramsci's famous phrase—but it is not continually deployed actively to control people's actions. To quote Adam Przeworski (1985, p. 136), “A hegemonic system is, for Gramsci, a capitalist society in which capitalists exploit with consent of the exploited.” For hegemony to be sustained over time, there must be, in Przeworski's (1985, pp. 133–69) apt expression, “material bases of consent.” This, in turn, requires some sort of class compromise: “the fact of hegemony presupposes that account be taken of the interests and the tendencies of the groups over which hegemony is to be exercised, and that a certain compromise equilibrium should

⁵ Although the actual term “class compromise” appears mainly within the Marxian tradition of social theory, the substantive idea has much broader currency. I will not limit the discussion here to instances where the term is explicitly deployed.

		Form of Class Compromise	
		NEGATIVE	POSITIVE
Strategic Basis of Class Compromise	INDIVIDUAL STRATEGIES WITHOUT ASSOCIATIONAL POWER	<i>Effort-promoting efficiency wages</i>	<i>Loyalty-promoting internal labor markets</i>
	ASSOCIATIONAL POWER	<i>Collective bargaining as mutual concessions</i>	<i>Positive-sum social pacts: Keynesianism; neocorporatism</i>

FIG. 3.—The conceptual space of class compromise

be formed—in other words, that the leading group should make sacrifices of an economic-corporate kind” (Gramsci 1971, p. 161).

Gramsci developed the concept of class compromise in only a sketchy and fragmented form. The scholar who has most systematically and rigorously elaborated this concept is Przeworski. Przeworski makes the central quid pro quo of class compromise explicit: “Given the uncertainty whether and how capitalists would invest profits, any class compromise must consist of the following elements: workers consent to profit as an institution, that is, they behave in such a manner as to make positive rates of profit possible; and capitalists commit themselves to some rate of transformation of profits into wage increases and some rate of investment out of profits” (Przeworski 1985, p. 182).

Przeworski’s formulation here is close to what I have called “negative class compromise,” insofar as he emphasizes the abstention of workers from levels of militancy that would interfere with the production of profits in exchange for material concessions by capitalists. Elsewhere, he explores the positive face of class compromise in his analysis of how Keynesianism, backed by organized labor and social-democratic parties in the advanced capitalist countries in the post–World War II period, expanded aggregate demand in ways that ultimately benefited capitalists as well as laborers (Przeworski 1985, pp. 205–11). The model of class compromise that I develop in this article can be viewed as an extension and reformulation of Przeworski’s core idea through the elaboration of this positive side of class compromise.

Before examining the details of this model, it will be useful to situate it within a broader array of alternative treatments of the problem of cooperation and compromise among class actors. Figure 3 organizes this con-

ceptual space along two dimensions: first, whether the strategic basis of class compromise is primarily *individual strategies* or *associational power*, and second, whether the form of class compromise is primarily *negative* or *positive*. The four categories generated by these two dimensions constitute distinctive ways in which interclass cooperation and compromise can be generated.

The paradigmatic case of negative class compromise grounded in individual strategies is so-called "efficiency wages."⁶ An efficiency wage is a kind of "employment rent"—a wage premium above the equilibrium "market-clearing wage"—paid by an employer as part of a strategy to reduce shirking on the part of employees. As elaborated by Bowles (1985) and Bowles and Gintis (1990, 1998, pp. 36–39), building on the earlier work of Alchian and Demsetz (1972), Shapiro and Stiglitz (1984), Akerlof and Yellen (1986), Williamson (1985), and others, employers face a problem of the "extraction of labor effort" from workers—getting workers to work harder than they want to do spontaneously—since the labor contract is neither complete nor costlessly enforceable. Employers face a trade-off between spending more money on improving the effectiveness of monitoring or paying higher employment rents. Such efficiency wages are a form of negative class compromise insofar as the higher wages are an alternative to more purely coercive strategies by employers in the face of strategies of resistance (shirking) by individual workers.⁷

Positive class compromises can also emerge out of individual strategic interactions between employers and workers. Perhaps the best example is internal labor markets (ILMs).⁸ Although, as in the case of efficiency

⁶ There are some treatments of efficiency wages that treat them more as a form of *positive* than of *negative* class compromise. For example, Akerlof (1982) sees such arrangements less as a concession in response to a form of individual resistance—shirking—and more as a way of improving generalized morale through a kind of normatively grounded "gift exchange." The implication is that efficiency wages underwrite active cooperation.

⁷ Investigations of institutional arrangements like efficiency wages, especially by economists, generally do not explicitly analyze them in class terms. From the point of view of standard neoclassical economics, such arrangements are simply profit-maximizing strategies of employers designed to minimize the transaction costs associated with the inherent human tendency, in Williamson's (1985, p. 47) expression, for people to be "self-interest seeking with guile." Bowles and Gintis (1990, 1998), in contrast, firmly situate the problem of efficiency wages in the class relations of capitalist production, arguing that in firms where workers were also owners, mutual monitoring would significantly replace the need for efficiency wages.

⁸ Some discussions of internal labor markets treat them primarily as examples of negative class compromise, emphasizing the ways in which ILMs are instigated to divide the working class, weaken unions, and in other ways enhance capitalist control over labor (see, e.g., Gordon [1976]; Edwards [1979]). Others treat ILMs strictly as a question of solving problems of internal efficiency, typically linked to information costs and the problem of retaining skilled employees, without systematic reference to the

wages, internal labor markets may increase the effectiveness of negative sanctions by employers since workers within internal labor markets have more to lose if they are disciplined or fired (Bartlett 1989, pp. 135–37), most analyses of internal labor markets emphasize the ways they are designed to elicit active cooperation rooted in loyalty and commitment of the individual to the interests of the organization. This is one of the central themes in the extensive literature on Japanese work organization (Dore 1973; Ouchi 1981; Aoki 1988), but it has also figured in the broader analysis of the internal organization of capitalist firms (Williamson 1985; Foulkes 1980; Sorensen 1994) and even in some more abstract analyses of class relations as such.⁹

Most discussions of class conflict and class compromise pay relatively little attention to these forms of interclass compromise, positive or negative, generated by the strategies of individuals. Rather, they focus on the ways class compromises are forged through class struggles rooted in class-based associational power. Analyses of negative class compromise emerging from class struggle are particularly prominent in the Marxist tradition.¹⁰ If the interests of workers and capitalists are inherently antagonistic and polarized, then it would appear that whatever compromises emerge from class struggle simply reflect balances of power between contending forces (for an illustrative example, see Kotz 1994, p. 55).

The idea of positive class compromise generated by organized class struggle sits less comfortably within the Marxist tradition but is at the core of the large literature on social democracy and neocorporatism (e.g., Korpi 1983; Soskice 1990; Esping-Andersen 1990) and considerable recent work in economic sociology that focuses on the problem of the economic

class character of these efficiency considerations (see, e.g., Doeringer and Piore 1985; Greenwald 1979; Waldman 1984).

⁹ Goldthorpe's (1982) concept of the "service class" revolves around the problems employers face when their employees sell a "service" rather than simply "labor." The creation of career ladders built around prospective rewards that create a longer time horizon of commitment for such employees is at the core of his analysis of the specificity of this employment relation. My analysis of "loyalty rents" for managerial class locations (Wright 1997, p. 21) also emphasizes the problem of creating deeper commitments for certain categories of employees by anchoring their jobs in career ladders. In both of these treatments of class relations, internal labor markets are created as responses to the strategies of individuals within firms.

¹⁰ Such views, however, are not restricted to Marxists. John R. Commons's ([1950] 1970) conception of collective bargaining, for example, is essentially a conception of negative class compromise insofar as he felt it was necessary to avoid a mutually destructive "class war." Unions might be in the general "public interest," but Commons does not claim that unionization and collective bargaining as such are directly beneficial to capitalists: "The unions and administrative commissions were organized to restrain corporations, also in the public interest, from abuse of their corporate power over individuals" (Commons 1970, p. 132).

performance of different capitalist economies (e.g., Streeck and Schmitter 1985; Kenworthy 1995; Gordon 1996; Crouch and Streeck 1997). As Rogers and Streeck (1994, p. 130) put it: "The democratic left makes progress under capitalism when it improves the material well-being of workers, solves a problem for capitalists that capitalists cannot solve for themselves, and in doing both wins sufficient political cachet to contest capitalist monopoly on articulating the 'general interest.'"

The classic form of this argument is rooted in the Keynesian strand of macroeconomic theory. Full employment, insofar as it implies high levels of capacity utilization and higher aggregate demand for the products of capitalist firms, potentially serves the interests of capitalists. But it also risks a profit squeeze from rapidly rising wages and spiraling levels of inflation. Keynes himself recognized this as a serious problem: "I do not doubt that a serious problem will arise as to how wages are to be restrained when we have a combination of collective bargaining and full employment" (cited in Glynn 1995, p. 37). The emergence and consolidation in a number of countries of strong, centralized unions capable of imposing wage restraint on both workers and employers was perhaps the most successful solution to this problem. In this sense, a powerful labor movement need not simply constitute the basis for a negative class compromise, extracting benefits for workers through threats to capital. If a labor movement is sufficiently disciplined, particularly when it is articulated to a sympathetic state, it can positively contribute to the realization of capitalists' interests by helping to solve macroeconomic problems.

The best-known empirical study to explore the curvilinear relationship between workers' power and capitalists' interests is Calmfors and Driffill's (1988) study of the effects of union centralization on economic performance (see also Pohjola 1992; Freeman 1988; Calmfors 1993; Garrett 1998, pp. 26–50; Rowthorn 1992).¹¹ Following Mancur Olson's (1982) original idea, Calmfors and Driffill (1988, p. 15) argue that "organized interests may be most harmful when they are strong enough to cause major disruptions but not sufficiently encompassing to bear any significant fraction of the costs for society of their actions in their own interests." They demonstrate that among 18 OECD countries, during the period 1963–85, economic performance measured in a variety of ways was best among those countries with *either* highly centralized *or* highly decentralized wage bargaining structures, and worst in the intermediary countries. A similar result, using different kinds of indicators, is found in Hicks and Kenwor-

¹¹ Strictly speaking, Calmfors and Driffill (1988) study the relationship between workers' power and various measures of general economic "performance" rather than capitalists' interests as such, but in the context of their arguments, this can reasonably be taken as an indicator of capitalists' interests.

thy's (1998) study of the impact of various forms of cooperative institutions on economic performance. They observe a strong curvilinear relationship between union density and real per capita gross domestic product growth for the period 1960–89 in 18 OECD countries, indicating that countries with either low or high union density had higher growth rates during these three decades than countries with middling levels of union density.

The rest of this article will attempt to elaborate theoretically this curvilinear model of positive class compromise. The next section begins by framing the problem through a game-theoretic perspective on strategic conflicts between workers and capitalists and then turns to the problem of the mechanisms that generate the curvilinear relationship between workers' power and capitalists' interests.

STRATEGIC GAMES AND CLASS COMPROMISE

In order to analyze the relationship of working-class associational power to capitalist-class interests and class compromise, we must first more rigorously understand the strategic contexts for the conflicts of interests of workers and capitalists. We will do this by exploring a series of stripped-down, game-theory models based on a highly simplified picture of class conflict in which workers and capitalists each face a binary strategic choice: to *cooperate* with the other class or to actively *oppose* its interests. Because the actors in this game have qualitatively different roles in the system of production, the meaning of "cooperate" and "oppose" are different for each. As summarized in figure 4, for workers to cooperate with capitalists means that they work hard and diligently in order to maximize the capitalists' rate of profit. Workers rely primarily on market mechanisms (changing jobs) as a way of expressing dissatisfaction with pay or working conditions; while they may have collective associations (unions), they do not engage in active struggles to collectively pressure capitalists for improvements; nor do they engage in political struggle to advance workers' interests against those of capitalists. To oppose capitalists is to struggle against them, individually and collectively, in order to raise worker incomes and enhance the extent to which workers control their own labor effort, and thus to minimize the extent to which capitalists exploit and control workers. This includes political struggles to expand worker rights and their capacity to organize collective associations. For capitalists, cooperation with workers means paying workers as much as is possible, compatible with maintaining a rate of profit sufficient to reproduce the firm; accepting workers' organizations (unions and parties) and responding positively to worker demands over working conditions; and moderating their own consumption in favor of employment-generating

		<i>Workers</i>	
		Cooperate with capitalists	Oppose capitalists
<i>Capitalists</i>	Cooperate with workers	C,C	O,C
	Oppose workers	C,O	O,O

Meaning of Cooperate & Oppose for Different Classes	
<i>Workers:</i>	
Cooperate:	Work hard and diligently to maximize the profits of capitalists
Oppose:	Minimize the extent to which capitalists exploit and dominate workers
<i>Capitalists</i>	
Cooperate:	Pay workers as much as possible and treat them fairly in the work place; accept workers' organizations
Oppose:	Pay workers as little as possible and get as much work out of them as possible; resist workers' organizations

FIG. 4.—Strategic options for workers and capitalists

investment. To oppose workers' interests means paying them as little as possible, given market and technological constraints; getting as much labor as possible out of workers; and resisting worker organizations. As in the case of workers, such opposition includes political action such as opposing unemployment benefits and welfare safety nets, which raise the reservation wage, and supporting restrictive labor laws that impede unionization. Taking these two alternatives for each class yields the four possible configurations of class conflict presented in figure 4. In terms of

these alternatives, “positive class compromise” constitutes the situation in which both classes agree to cooperate (C,C).¹²

Figure 5 presents a variety of alternative ways in which the interests of workers and capitalists (their payoffs to alternative strategic combinations) might be affected by these four combinations of cooperation and opposition. Model 1 can be called a unilateral capitalist domination game. Here, the best outcome for capitalists is (C,O): workers cooperate with capitalists (working hard, not organizing, etc.), and capitalists oppose workers (pay them only what the market dictates, oppose collective organization, etc.). The second best outcome for capitalists is mutual opposition (O,O). In this game, capitalists are sufficiently powerful relative to workers that they can punish workers at relatively little cost to themselves when workers organize against them. Workers are thus worse off under (O,O) than under *unilateral* workers cooperation (C,O). Struggle does not pay. In this game, therefore, (C,O) will be the equilibrium outcome: capitalists are always better off opposing workers, and given that capitalists oppose workers, workers are better off cooperating with capitalists.

Model 2 represents the standard Marxist view of class conflict in which the interests of workers and capitalists are treated in a purely inverse relation as a zero-sum pure conflict game. The optimal situation for capitalists is that they oppose the interests of workers while workers cooperate with them (C,O). The second best situation for capitalists is mutual cooperation (C,C). This, however, is *less* advantageous for workers than is mutual opposition (O,O). In the traditional Marxist view, because the interests of workers and capitalists are strictly polarized, it is always better for workers to struggle against capitalists—to actively oppose capitalists’ interests—than to willingly cooperate. The (C,C) solution, in effect, is an illusion: “cooperative” capitalists, the argument goes, treat workers only marginally better than capitalists who actively oppose workers, but cooperative workers are much less able to force their employers to make concessions than are oppositional workers. Above all, when working-class associations actively cooperate with capitalists, they weaken their capacity for mobilization, and ultimately this invites capitalists to oppose workers’ interests, thus leading (C,C) to degenerate into (C,O). The (O,O) option, therefore, generally promises a better long-term payoff for workers

¹² Obviously in the real world, the options are much more complex than this stark contrast—not only are there various degrees of opposition and cooperation, but a variety of qualitatively distinct forms of both. Nevertheless, for purposes of developing a general inventory of strategic contexts for class compromise in which mutual cooperation occurs, it will be useful to abstract from such complexity and examine games in which members of each class (considered either as individuals or as members of associations) make such simple, dichotomous choices.

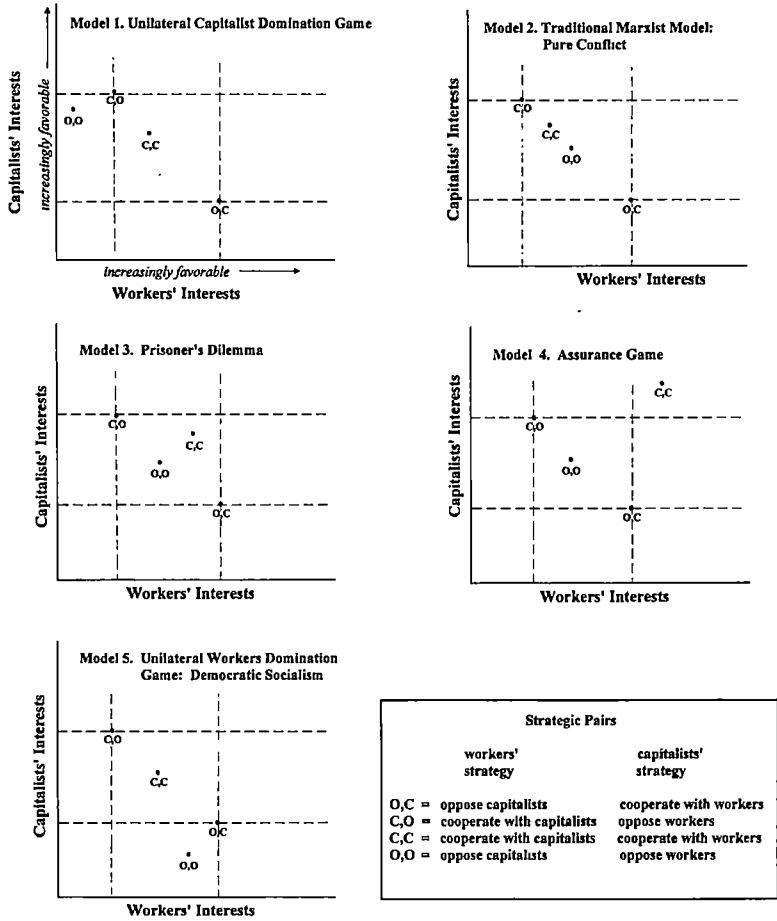


FIG. 5.—Possible strategic games and payoffs for workers and capitalists

than does the (C,C) option. As a result of such struggles, there will be moments when capitalists indeed do make concessions to workers as a result of these struggles—grant them pay raises, improve working conditions, and so on. These concessions are at best a negative class compromise—concessions in the face of struggle. For both classes in this game, opposition is better than cooperation regardless of what the other class does, and thus the equilibrium will be mutual opposition (O,O)—active forms of class struggle.¹³ The class struggle is much more like trench war-

¹³ Even though mutual opposition is the equilibrium solution, model 2 is not a prisoner's dilemma for workers since in a prisoner's dilemma actors prefer mutual cooperation to mutual opposition.

fare with occasional victories and defeats for each combatant, and perhaps periods of relatively stable balances of forces underwriting a *negative* class compromise.

Model 3 is the standard prisoner's dilemma game. This is a game with symmetrical payoffs for the two classes: (C,C) is the second-best outcome for each class, and (O,O) is the third-best outcome. Unlike in model 1, mutual opposition is now costly to capitalists. This implies that workers have sufficient power to be able to punish capitalists within class struggles. Unlike in model 2, however, workers are better off in mutual cooperation than in mutual opposition. Both classes are thus better off if they cooperate with each other than if they mutually oppose each other. Still, if this were a one-shot game, in standard PD fashion, the equilibrium outcome would be (O,O), since both classes could improve their payoffs by defecting from the mutual cooperation outcome. If this is a repeated game, as it would be in the real world of class interactions, then the outcome is less determinate. As Axelrod (1984) and many others have shown, in an iterated prisoner's dilemma, mutual cooperation can be a stable solution depending upon the ways opposition in future rounds of the game is used to punish players for noncooperation in earlier rounds. As the possibility of a stable (C,C) solution occurs, then *positive* class compromise also becomes possible.

Model 4 is a standard assurance game: for both classes, the optimal solution is mutual cooperation and unilateral cooperation is worse than mutual opposition. Unless there is reasonable confidence that the other class will cooperate, therefore, mutual cooperation is unlikely to occur. If class conflict was an assurance game, the failures of cooperation would primarily reflect a lack of enlightenment on the part of actors—they simply do not know what is good for them. This corresponds to the views of a certain kind of naive liberalism, where conflict is always seen as reflecting misunderstanding among parties and “win-win” solutions are always assumed to be possible.

A strict assurance game of this form is unlikely in capitalist economies since a situation in which capitalists can get full cooperation from workers without having to make any concessions—the (C,O) outcome—is unlikely to offer capitalists inferior payoffs to mutual cooperation. Nevertheless, there may be situations in which the (C,C) payoff moves in the direction of an assurance game and certainly situations in which the gap for both classes between (C,C) and (O,O) becomes very large.

Finally, model 5, the unilateral workers domination game, is the symmetrical model to model 1. Here workers are sufficiently strong and capitalists are sufficiently weak that workers can force capitalists to unilaterally cooperate, including forcing them to invest in ways that enhance future earnings of workers (thus making [O,C] preferable to [C,C] for

workers). This corresponds to the theoretical idea of democratic socialism: an economy within which workers effectively dominate capitalists.¹⁴

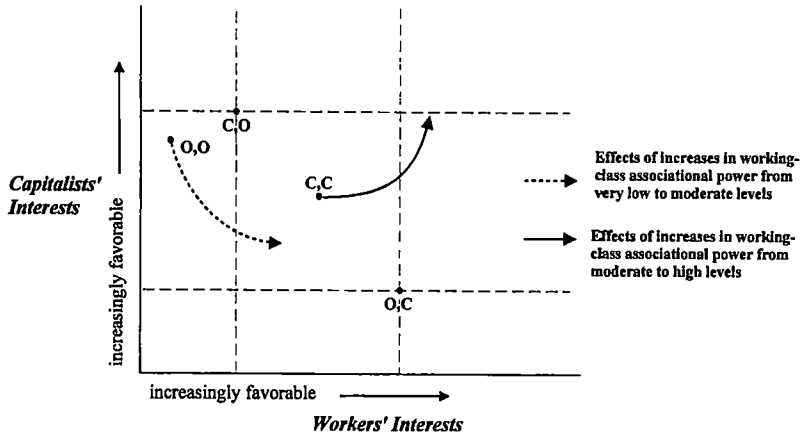
Lurking in the background of the models in figure 5 is the problem of power: the balance of power between workers and capitalists can be thought of as determining which of these strategic games is being played. As illustrated in figure 6, as working-class power increases from extremely low levels (and thus as the ability of workers to impose sanctions on capitalists increases), the (O,O) alternative in model 1 shifts downward and then to the right. This shifts the configuration in the direction of model 2 in which working-class militancy becomes sustainable and the possibility of negative class compromise—a class compromise based on the balance of force—emerges. Further increases in working-class power begin to move the (C,C) option to the right, creating the prisoner's dilemma of model 3. This sets the stage for the possibility of *positive* class compromise. As working-class associational power pushes (C,C) in model 3 in an upward direction toward the northeast quadrant approaching the assurance game in model 4, the possibility for a *positive* class compromise increases: the gains from stable, mutual cooperation increase.¹⁵

If it were the case that increases in working-class associational power could actually push the (C,C) payoff into the northeast quadrant of this payoff matrix, then the overall relationship between workers' power and capitalists' interests would be a J-curve, not a reverse-J. That is, the highest across-game equilibrium payoff for capitalists would be the (C,C) payoff in the assurance game rather than the (C,O) payoff in the unilateral capitalist domination game, and thus it would be in the interests of capitalists to accept (and even encourage) high levels of workers' power in order to create the conditions for the assurance game to occur. It is a central substantive assumption of the Marxian framework deployed in this article that because of the underlying antagonistic, exploitative character of capitalist-class relations, this situation does not occur. Stable, mutual cooperation can still occur, but it is because, with sufficient power, the threat of opposition by workers prevents the (C,O) option from being an equilib-

¹⁴ In this theoretical conception of socialism, capitalists, albeit with curtailed property rights, can exist within a socialist economy just as they existed centuries earlier within feudal society. It is another question how stable and reproducible such a structure of class relations would be. For a formal model of a sustainable socialist society within which capitalists still have some economic space, see Roemer (1994, 1996).

¹⁵ As portrayed in fig. 6, working-class power only affects the (O,O) and (C,C) curves; the (C,O) and (O,C) curves remain fixed. The critical issue in the shifts across the models in fig. 5, of course, is the change in the *relative* location of the four payoffs, and this in principle could occur because of changes in the location of (C,O) or (O,C) as well as the mutual opposition or mutual cooperation payoffs.

Transformation of the *Unilateral Capitalist Domination Game*
into a *Prisoner's Dilemma* and then toward an *Assurance Game*



Working-Class Power and the Transformation
of the O,O and C,C Payoffs for Capitalists

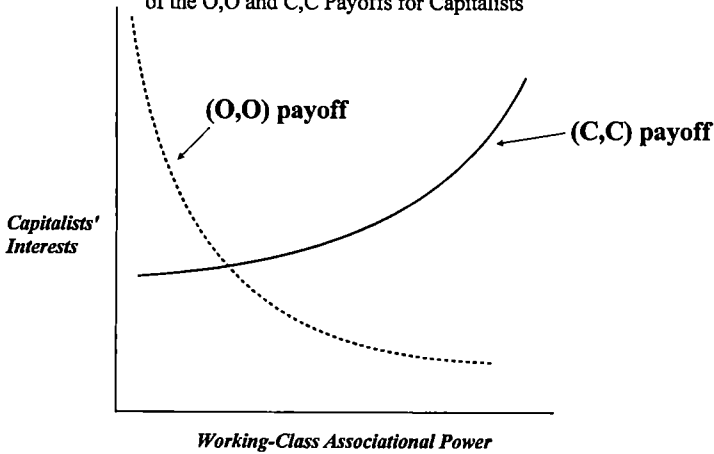


FIG. 6.—Working-class power and the transformations of the strategic context of class compromise.

rium, not because mutual cooperation is the best of all possible payoffs for capitalists.¹⁶ The two curves in figure 6 are both nonlinear: one, (O,O), downward sloping and the other, (C,C), upward sloping. The nonlinear shape of these relations is important for the proposed reverse-J model of class compromise, since if these two curves were each linear, they would generate a linear overall relation between workers' power and capitalists' interests. The shape of the downward sloping (O,O) curve occurs because relatively modest levels of workers' power can create considerable damage to capitalists' interests but are insufficient to generate much sustainable gain for workers. Increases in workers' power from negligible to moderate, therefore, increase the punishment capacity of workers considerably. Beyond a certain point, however, there are diminishing returns in the additional degree of harm to capitalists generated by additional working-class strength. Once workers are sufficiently strong to prevent capitalists from arbitrarily firing workers, for example, being even stronger does not yield additional gains in job security. The (C,C) curve is nonlinear and upward sloping because the positive gains capitalists can realize by virtue of workers' power only occur when workers are sufficiently well organized and solidaristic that their associations can effectively sanction defectors from cooperation both among their own members and among capitalists. Until worker associations are at least moderately powerful, they lack this dual-disciplining capacity and thus generate little positive effect on capitalists' interests.

This, then, is the central game-theoretic logic underlying the argument developed in this article: as working-class power increases, the unilateral capitalist domination game is initially shifted to a pure conflict game, making negative class compromise possible; with further increases in working-class associational strength, the strategic environment can shift toward an iterated prisoner's dilemma, opening the prospect for positive class compromise. The more the game shifts toward an assurance game—even though it is unlikely to actually become one—the more stable the possibility of positive class compromise will become. Underlying this double shift is thus the problem of the relation of working-class associational power to the interests of capitalists, to which we now turn.

¹⁶ It is difficult to find direct empirical evidence that the shape of the curve is a reverse-J, i.e., that capitalists are best off in the (C,O) equilibrium of the unilateral capitalist domination game. The observation that in countries with relatively disorganized working classes, such as the United States, the capitalist class and CEOs are personally much richer than in countries with highly organized working classes is consistent with the reverse-J argument, but it is potentially confounded by many other factors.

A CURVILINEAR MODEL OF POSITIVE CLASS COMPROMISE

To the extent that increases in working-class power can contribute not merely to the realization of working-class material interests, but also to the realization of some capitalist-class interests, class compromises are likely to be more stable and beneficial for workers. To the extent that every increase in working-class power poses an increasing threat to capitalist-class interests, capitalists' resistance is likely to be more intense, and class compromises, even if achieved, are likely to be less stable. The intensity of class struggle, therefore, is not simply a function of the relative balance of power of different classes, but also of the intensity of the threat posed to dominant interests by subordinate-class power.

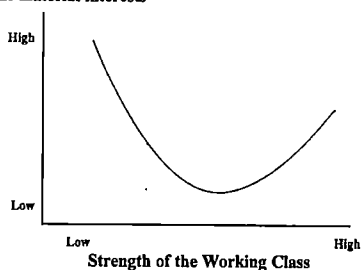
If the relationship between workers' power and capitalists' interests were the simple inverse relationship of figure 1, then class compromises would always be relatively fragile and vulnerable to attack, for capitalists' interests would always be served by taking advantage of opportunities to undermine workers' power. Negative class compromise would be the most one could achieve. If the shape of the relationship is as pictured in figure 2, on the other hand, then class compromise can potentially become a relatively durable feature of a set of institutional arrangements. In general, when class conflict is located in the upward-sloping region of this curve, class compromises are likely to be both more stable and more favorable for the working class. If the shape of this curve assumes the form of a more U-shaped version of a reverse-J (i.e., if the upward-sloping section becomes more symmetrical), then conditions for class compromise can be said to be more favorable; if the reverse-J degenerates into a strictly downward-sloping curve, then the conditions for class compromise become less favorable.

In order to more deeply understand the social processes reflected in the reverse-J hypothesis of figure 2, we need to elaborate and extend the model in various ways. First, we will examine more closely the underlying causal mechanisms that generate this curve. Second, we will extend the range of the figure by examining what happens at extreme values of working-class associational power. Finally, we will examine various ways in which the institutional environment of class conflict determines which regions of this curve are historically accessible as strategic objectives.

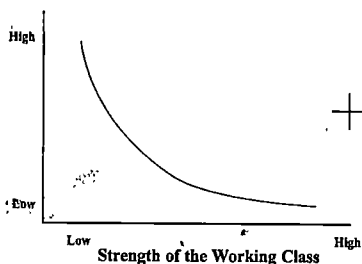
Mechanisms Underlying the Reverse-J Relation

The reverse-J curve presented in figure 2 can be understood as the outcome of two kinds of causal processes—one in which the interests of capitalists are increasingly undermined as the power of workers increases and a second in which the interests of capitalists are enhanced by the increas-

Capacity of capitalists
to realize material interests



Capacity of capitalists to
unilaterally make decisions
and control resources



Capacity of capitalists
to solve collective action
and coordination problems

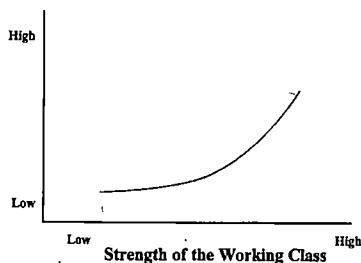


FIG. 7.—Decomposition of the relationship between interests of capitalists and associational strength of workers.

ing power of workers. These are illustrated in figure 7. In broad terms, the downward-sloping curve reflects the ways in which increasing power of workers *undermines the capacity of capitalists to unilaterally make decisions and control resources* of various sorts, while the upward-sloping curve reflects ways in which the associational power of workers may *help capitalists solve certain kinds of collective action and coordination problems*. The specific nonlinear shapes of these curves are derived from the shapes of the curves in figure 6.

The mechanisms that generate the component curves in figure 7 can be differentiated across the three institutional spheres within which class compromises are forged: exchange, production, and politics. These mechanisms are summarized in figure 8.

The sphere of exchange.—Capitalists have a range of material interests within the sphere of exchange that bear on their relationship with the working class: minimizing labor costs, having an unfettered capacity to hire and fire without interference, selling all of the commodities they pro-

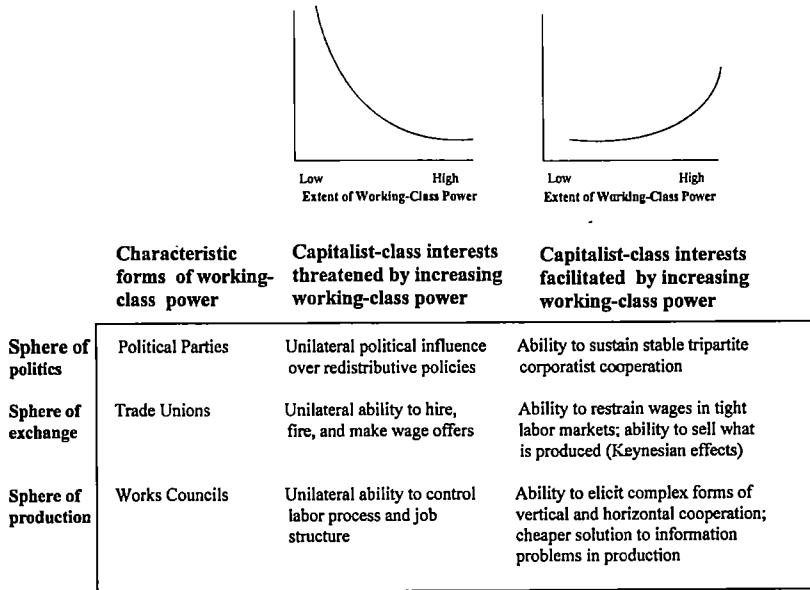


FIG. 8.—Decomposition of the relationship between working-class power and capitalist-class interests in the sphere of politics, exchange, and production.

duce, and having a labor force with a particular mix of skills in a labor market that provides predictable and adequate supplies of labor. As has often been argued by both Marxists and non-Marxist political economists, some of these interests contradict each other. Most notably, the interest of capitalists in selling commodities means that it is desirable for workers-as-consumers to have a lot of disposable income, whereas capitalists' interest in minimizing their own wage bill implies an interest in paying workers-as-employees as little as possible.

Increases in working-class associational power generally undermine the capacity of individual capitalists to unilaterally make decisions and allocate resources within labor markets. In the absence of unions, capitalists can hire and fire at will and set wages at whatever level they feel is most profitable. Of course, this does not mean that employers set wages without any constraints whatsoever. Their wage offers will be constrained by the tightness or looseness of the labor market, the reservation wages of workers, and, as discussed earlier, the need to pay workers a sufficiently high wage to motivate individual workers to work diligently. Capitalists' decisions are thus always constrained by the actions of individual workers and by general economic conditions. The issue here, however, is the extent of constraint on capitalists imposed by the collective action of workers

reflecting their associational power in various forms. Such associational power reduces capitalists' individual capacity to make profit-maximizing decisions on labor markets and thus hurts their material interests.

If capitalists' interests within the sphere of exchange consisted entirely of interests in their individual ability to buy and sell with minimal constraint, then something close to the inverse relation portrayed in figure 1 would probably hold. But this is not the case. The material interests of capitalists—their ability to sustain a high and stable rate of profit—depends upon the provision of various aggregate conditions within the sphere of exchange, and these require coordination and collective action. The solution to at least some of these coordination problems can be facilitated by relatively high levels of working-class associational power.¹⁷

The classic example of this is the problem of inadequate aggregate demand for the consumer goods produced by capitalists. This is the traditional Keynesian problem of how raising wages and social spending can underwrite higher levels of aggregate demand and thus help solve "under-consumption" problems in the economy. Inadequate consumer demand represents a collective action problem for capitalists: capitalists simultaneously want to pay their own employees the lowest wages possible and want other capitalists to pay the highest wages possible in order to generate adequate consumer demand for products. High levels of unionization, in effect, prevent individual firms from "defecting" from the cooperative solution to this dilemma. Working-class strength can also contribute to more predictable and stable labor markets. Under conditions of tight labor markets, where competition for labor among capitalists would normally push wages up, perhaps at rates higher than the rate of increase of productivity, thus stimulating inflation, high levels of working-class associational power can also contribute to wage restraint (see Calmfors and Driffill 1988; Glynn 1995; Pontusson 1997). Wage restraint is an especially complex collective action problem: individual capitalists need to be prevented from defecting from the wage-restraint agreement (i.e., they must be prevented from bidding up wages to workers in an effort to lure workers away from other employers given the unavailability of workers in the labor market), and individual workers (and unions) need to be prevented from defecting from the agreement by trying to maximize wages under tight labor market conditions. Wage restraint in tight labor markets,

¹⁷ This does not mean that working-class associational power is a necessary condition for the solution to such coordination problems. There may be other devices that may constitute alternative strategies for accomplishing this. All that is being claimed is that working-class associational power can constitute a mechanism that makes it easier to solve such problems.

which is important for longer term growth and contained inflation, is generally easier where the working class is very well organized, particularly in centralized unions, than where it is not.

These positive effects of workers' strength on capitalists' interests in the sphere of exchange need not imply that capitalists themselves are equally well organized in strong employers associations, although as the history of Northern European neocorporatism suggests, strongly organized working-class movements tend to stimulate the development of complementary organization on the part of employers. In any case, the ability of workers' power to constructively help solve macroeconomic problems is enhanced when capitalists are also organized.

Assuming that the positive Keynesian and labor market effects of working-class power are generally weaker than the negative wage cost and firing discretion effects, the combination of these processes yields the reverse-J relationship for the sphere of exchange in figure 8.

The sphere of production.—A similar contradictory quality of the interests of capitalists with respect to workers occurs within the sphere of production: on the one hand, capitalists have interests in being able to unilaterally control the labor process (choosing and changing technology, assigning labor to different tasks, changing the pace of work, etc.), and on the other hand, they have interests in being able to reliably elicit cooperation, initiative, and responsibility from employees.

As working-class associational power within production increases, capitalists' unilateral control over the labor process declines. This does not mean that capitalists are necessarily faced with rigid, unalterable work rules, job classifications, and the like, but it does mean that changes in the labor process need to be negotiated and bargained with representatives of workers rather than unilaterally imposed. Particularly in conditions of rapid technical change, this may hurt capitalists' interests.

On the other hand, at least under certain social and technical conditions of production, working-class associational strength within production may enhance the possibilities for more complex and stable forms of cooperation between labor and management. To the extent that working-class strength increases job security and reduces arbitrariness in managerial treatment of workers, the workers' time horizons for their jobs are likely to increase and along with this their sense that their future prospects are linked to the welfare of the firm. This in turn may contribute to a sense of loyalty and greater willingness to cooperate in various ways.

The German case of strong workplace-based worker organization built around works councils and codetermination is perhaps the best example. Streeck describes how codetermination and works councils positively help capitalists solve certain problems:



What, then, is specific about codetermination? Unlike the other factors that have limited the variability of employment, codetermination has not merely posed a problem for enterprises, but has also offered a solution. While on the one hand codetermination has contributed to growing organizational rigidities, on the other hand, and at the same time, it has provided the organizational instruments to cope with such rigidities without major losses in efficiency. . . . The works councils not only shares in what used to be managerial prerogatives, but also accepts responsibility for the implementation and enforcement of decisions made under its participation. This constellation has frequently been described as "integration" or "cooptation" of labor or organized labor, in management; with the same justification, however it can be seen as "colonization" of management, and in particularly manpower management, by the representatives of the workforce. The most adequate metaphor would probably be that of a *mutual incorporation of capital and labor* by which labor internalizes the interests of capital just as capital internalizes those of labor, with the result that works council and management become subsystems of an integrated, internally differentiated system of industrial government which increasingly supersedes the traditional pluralist-adversarial system of industrial relations. (Streeck 1992, pp. 160, 164; emphasis in the original)

This tighter coupling of interests of labor and capital with the resulting heightened forms of interclass cooperation helps employers solve a range of concrete coordination problems in workplaces: more efficient information flows within production (since workers have more access to managerial information and have less incentive to withhold information as part of a job-protection strategy); more efficient adjustments of the labor process in periods of rapid technological change (since workers are involved in the decision making and are thus less worried that technological change will cost them their jobs, they are more likely to actively cooperate with the introduction of new technologies); and more effective strategies of skill formation (since workers, with the most intimate knowledge of skill bottlenecks and requirements, are involved in designing training programs). Most broadly, strong workplace associational power of workers creates the possibility of more effective involvement of workers in various forms of creative problem solving.¹⁸

With so many positive advantages of such cooperative institutions, it might seem surprising that strong workplace associational power is so rare

¹⁸ It is possible, under certain social and cultural conditions, for some of these forms of cooperation to emerge and be sustained without strong workplace associational power of workers. This is often the way the relatively cooperative system of employment relations in Japan is described (see, e.g., Nakane 1970), although others have criticized such culturalist views (e.g., Aoki 1988, pp. 304–13). In any event, under many conditions high levels of worker cooperation within production are likely to be difficult to sustain if they are not backed by some form of significant associational power.

in developed capitalist countries. The reason, as I have argued throughout this article, is that such cooperative advantages come at a cost to capital. Streeck recognizes this even in the German case: "Above all, codetermination carries with it considerable costs in managerial discretion and managerial prerogatives. . . . Integration cuts both ways, and if it is to be effective with regards to labor it must bind capital as well. This is why codetermination, for all its advantages, is seen by capital as a thoroughly mixed blessing. . . . Both the short-term economic costs and the long-term costs in authority and status make the advantages of codetermination expensive for the capitalist class, and thus explains the otherwise incomprehensible resistance of business to any extension of codetermination rights" (Streeck 1992, p. 165).

Because of these costs, capitalists in general will prefer a system of production in which they do not have to contend with strong associational power of workers in production. Thus, the reverse-J shape of the functional relation between workers' power and capitalists' interests within production is formed.

The sphere of politics.—The two components of the reverse-J relationship between working-class associational power and capitalists' interests are perhaps most obvious in the sphere of politics. As a great deal of comparative historical research has indicated, as working-class political power increases, the capitalist state tends to become more redistributive: the social wage increases and thus the reservation wage of workers is higher; taxation and transfer policies reduce income inequality; and in various ways, labor power is partially decommodified. All of these policies have negative effects on the material interests of high-income people in general and capitalists in particular. Working-class political power also tends to underwrite institutional arrangements that increase working-class power within the sphere of exchange and often within the sphere of production as well. Working-class associational power in the political sphere, therefore, may also contribute to the downward-sloping curves in the spheres of exchange and production.

The upward-sloping class compromise curve in the sphere of politics is the central preoccupation of social democracy. The large literature on tripartite state-centered corporatism is, in effect, a literature on how the interests of capitalists can flourish in the context of a highly organized working class (Esping-Andersen 1990; Schmitter and Lembruch 1979; Schmitter 1988). Sweden, until the mid-1980s, is usually taken as the paradigm case: the social-democratic party's control of the Swedish state facilitated a set of corporatist arrangements between centralized trade unions and centralized employers' associations that made possible a long, stable period of cooperation and growth. The organizational links between the labor movement and the social-democratic party were critical for this sta-

bility, since it added legitimacy to the deals that were struck and increased the confidence of workers that the terms of the agreement would be upheld in the future. This made it possible over a long period of time for Swedish capitalism to sustain high capacity utilization, very low levels of unemployment, and relatively high productivity growth. State-mediated corporatism anchored in working-class associational strength in the political sphere played a significant role in these outcomes.

The inventory of mechanisms in figure 8 provides a preliminary set of variables for characterizing the conditions of class compromise within different units of analysis across time and space. Class compromises within the sphere of exchange can occur in local, regional, or national labor markets, or within labor markets linked to particular sectors. Production-level compromises typically occur within firms, but they may also be organized within sectors.¹⁹ Class compromises in the sphere of politics are especially important within the nation state, but local and regional political class compromises are also possible. The emergence of various forms of mesocorporatism involving local and regional levels of government may indicate the development of political class compromises within subnational units. The reverse-J curves that map the terrain of class compromise, therefore, can be relevant to the analysis of class compromises in any unit of analysis, not simply entire countries.

Different countries, then, will be characterized by different combinations of values on these three pairs of class compromise curves.²⁰ In Ger-

¹⁹ In the spheres of production and exchange, there may be considerable heterogeneity in the shape of the class compromise curves and the degree of working-class associational power across firms and sectors. The result is that within a given country the conditions for class compromise may be much more favorable in some firms and sectors than in others. Within the sphere of production, it is easy enough to see how the upward-sloping curve can be restricted to a particular sector or even firm, since most of the gains from cooperation are contained within firms. In the sphere of exchange, while many of the positive effects of high levels of unionization for capitalists come from aggregate, macroeconomic effects, some of the positive effects—such as stabilization of labor markets, rationalized skill formation, and wage restraint in tight labor markets—may be concentrated in specific sectors or localities. The reverse-J curve characterizing a given sphere, therefore, is itself an amalgamation of the distribution of such curves across firms, sectors, and other less-aggregated units of analysis.

²⁰ The actual variation across time and place is, of course, much more complicated than is being portrayed here. Countries will vary not simply in where they are located on each of these curves, but also on: (1) the relative weights of the various curves in defining the overall configuration for the society; (2) the units of analysis within countries within which class compromises are most rooted; and (3) the specific shapes of the component curves themselves. In some times and places, for example, the upward-sloping segments of some of the curves might be relatively flat, and in other cases, quite steep. My theoretical understanding of these relations is insufficient to say anything very systematic about these sources of variation.

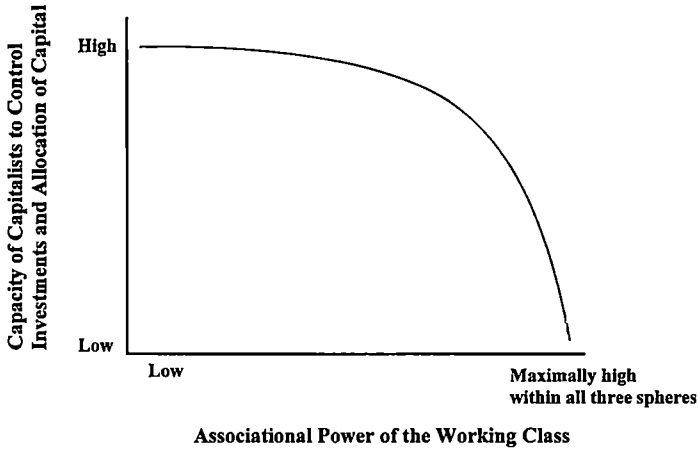


FIG. 9.—Interests of capital and power of workers with respect to the control over investments.

many, for example, working-class associational power is especially strong within the sphere of production, somewhat less strong in the sphere of exchange, and rather weaker in the sphere of politics. In Sweden—at least in the heyday of social democracy—it has been very strong in the spheres of exchange and politics and perhaps a bit weaker in the sphere of production. In the United States, working-class associational power has dwindled within all three spheres but is strongest in the sphere of exchange within certain limited sectors. The overall reverse-J curve for class compromise within a society, therefore, is the result of a complex amalgamation of the component curves within each of these spheres.

Making the Model More Complex: Extending the Theoretical Domain of Variation

The range of variation in figures 2 and 8 can be considered the typical spectrum of possibilities in contemporary, developed capitalist societies. It will be helpful for our subsequent analysis to consider what happens when working-class power increases toward the limiting case of society-wide, working-class organization and solidarity simultaneously in all three spheres of class compromise. This corresponds to what might be termed "democratic socialism," where socialism is not defined as centralized state ownership of the means of production but as working-class collective control over capital.

What happens to capitalist-class interests as working-class associational power approaches this theoretical maximum? Figure 9 presents the rela-

tionship between one crucial aspect of capitalists' interests—their control over investments and accumulation (allocation of capital)—and working-class power. The control over investments is perhaps the most fundamental dimension of "private" ownership of the means of production within capitalism. In most capitalist societies, even as working-class power increases, this particular power of capital is not seriously eroded. Even with strong unions and social-democratic parties, capitalists still have the broad power to disinvest, to choose their individual rate of savings, to turn their profits into consumption or allocate them to new investments, and so on. Of course, all capitalist states have capacities to create incentives and disincentives for particular allocations of capital (through taxes, subsidies, tariffs, etc.). And in special circumstances, "disincentives" can have a significant coercive character, effectively constraining capitalists' capacity to allocate capital. Still, this fundamental aspect of capitalists' property rights is not generally threatened within the normal range of variation of working-class power. When working-class associational power approaches its theoretical maximum, however, the right of capitalists to control the allocation of capital is called into question. Indeed, this is the heart of the definition of democratic socialism—popular, democratic control over the allocation of capital. This suggests the shape of the curve in figure 9: a relatively weak negative effect of working-class power on capitalists' interests with respect to the control over the basic allocation of capital until working-class power reaches a very high level, at which point those interests become seriously threatened.²¹

When figure 9 is added to figure 2, we get the roller-coaster curve in figure 10. There are two maxima in this theoretical model: the *capitalist utopia*, in which the working class is sufficiently atomized and disorganized to give capitalists a free hand in organizing production and appropriating the gains from increased productivity without fear of much collective resistance; and the *social-democratic utopia*, in which working-class associational power is sufficiently strong to generate high levels of corporatist cooperation between labor and capital without being so strong as

²¹ The x-axis in fig. 9 is working-class associational power undifferentiated into the spheres of production, exchange, and politics. It thus represents an undertheorized amalgam of the associational power within the three spheres (which are themselves amalgams of associational power across the various units of analysis that make up a sphere). The underlying intuition is that viable democratic socialism requires high levels of workers' associational power within all three spheres and that a *sustainable* threat to fundamental capitalists' property rights under democratic conditions can only occur when such unified associational power occurs. This does not imply, however, that the three spheres are of equal weight in this theoretical gestalt. Traditionally, Marxists have argued that working-class power at the level of the state is most decisive for challenging capitalist property rights, whereas syndicalists have argued that the pivot is workers' power within production.

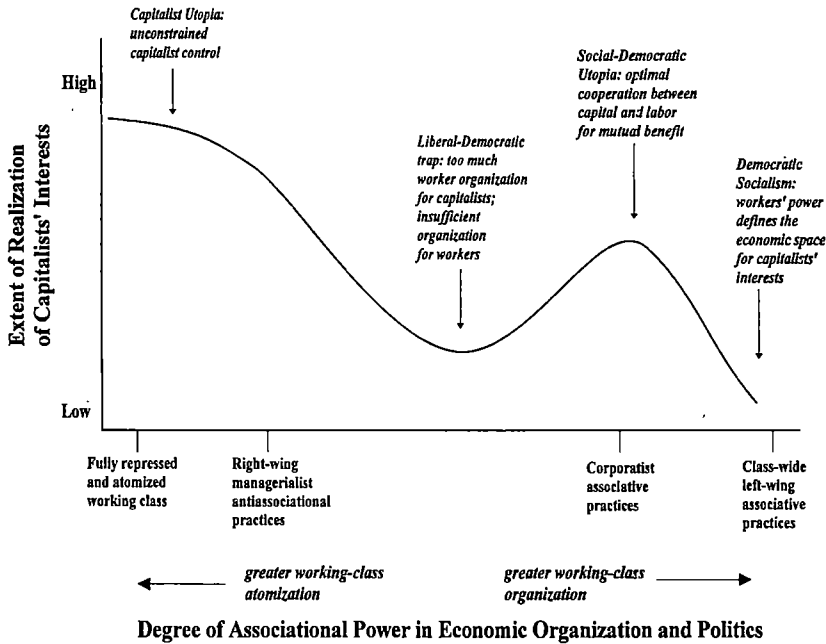


FIG. 10.—Expanded model of working-class associational power and capitalists' interests.

to threaten basic capitalist property rights. These two maxima, however, constitute quite different strategic environments for workers and capitalists. Statically, capitalists should only care about where they sit on the vertical axis of this figure: if you draw a horizontal line through the figure that intersects the curve at three places, capitalists should be statically indifferent among these three possibilities. Understood dynamically, however, capitalists in general will prefer points in the left-hand region of the curve.

It is at least in part because of this threat of a society-wide shift in the balance of class power that capitalists might prefer for working-class associational power to remain to the left of the social-democratic "peak" of this curve, even though this peak might be theoretically advantageous to capitalists' interests. Arriving at the peak looks too much like a Trojan Horse: small additional changes in associational power could precipitate a decisive challenge to capitalists' interests and power. The local maximum of the "social-democratic utopia" in figure 10 may thus be a kind of tipping point, which is seen by capitalists as too risky a zone to inhabit. This is one interpretation of the strident opposition by Swedish capitalists

to the initial formulation of the "wage-earners fund" proposal in Sweden in the 1970s. The wage-earners fund, as initially conceived, was a proposal through which Swedish unions would gain increasing control over the Swedish economy via the use of union pension funds to purchase controlling interests in Swedish firms. From the point of view of economic performance and even the middle-run profit interests of Swedish firms, it was arguable that this might be beneficial for Swedish capital, but it raised the possibility of a long-term slide toward democratic socialism by significantly enhancing the power of Swedish labor. The result was a militant attack by Swedish capitalists against the social-democratic party. As Glynn (1995, pp. 53–54) writes: "The policies which the Social Democrats were proposing impinged on the authority and freedom of action of business which was supposed to be guaranteed in return for full employment and the welfare state. This seems to lie at the root of the employers repudiation of the Swedish model, of which full employment was a central part."

The different regions of this curve correspond to the different game-theory models in figure 5. The capitalists' utopia corresponds to the unilateral capitalist domination game in which (C,O) is the equilibrium solution. The downward-sloping region in the center of the figure is the pure conflict game where, at best, negative class compromise is possible. The upward-sloping part of the curve is the iterated prisoner's dilemma, where a stable (C,C) solution, a positive class compromise, can emerge. The apex of this region of the curve, the social-democratic utopia, is the point that is closest to an assurance game. If in fact it actually became a proper assurance game (i.e., the [C,C] payoff in figure 5 moved into the northeast quadrant of the payoff matrix), then the central region of the curve in figure 10 would become a J-curve rather than a reverse-J; the social-democratic utopia would be higher than the capitalist utopia and become a kind of social-democratic nirvana in which mutual cooperation between classes was self-reinforcing, no longer resting on a background condition of potential working-class opposition. Finally, democratic socialism corresponds to the unilateral working-class domination game in which (O,C) is the equilibrium solution.

Working-Class Interests and the Class Compromise Curve

The models in figure 5 contain both working-class interests and capitalist-class interests. Figure 11 adds working-class interests to the class compromise curve in figure 10. The different regions of these curves can be thought of as specific hypotheses about the effects of marginal changes of working-class power on the relationship between workers' interests and capitalists' interests:

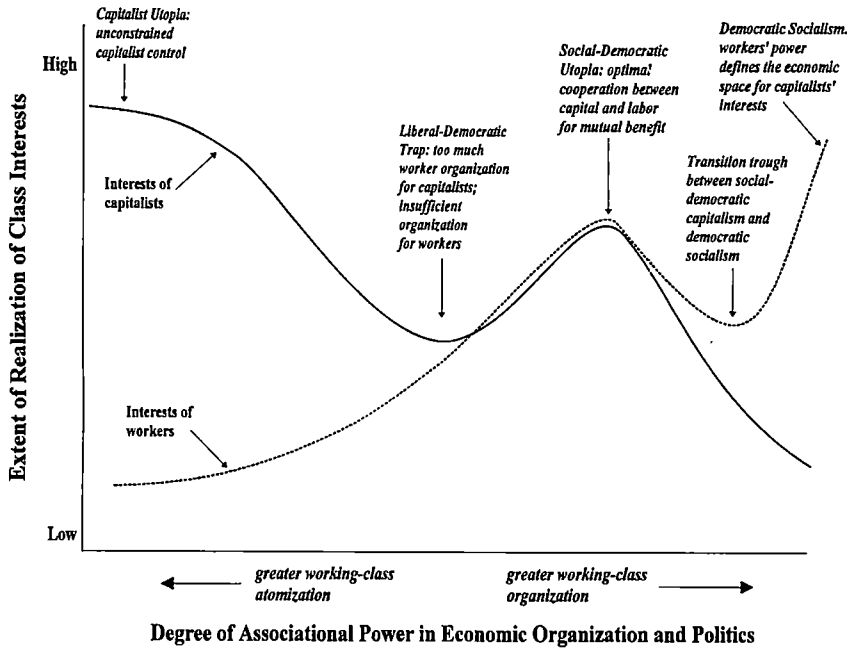


FIG. 11.—Working-class associational power, working-class interests, and capitalists' interests.

1. The gap between workers' interests and capitalists' interests is greatest at the ends of the spectrum: when working-class associational power is weakest (the fully atomized working class) or at the maximum strength (democratic socialism).
2. Increases in working-class associational power steadily increase the realization of working-class material interests up to relatively high levels of associational power. Of course, in actual historical processes of increasing working-class power, it may well happen that there will be episodes in which the resistance of capitalists results in declines in the realization of working-class interests. Nevertheless, in general, increasing workers' power is expected to improve the realization of working-class interests.
3. The region of the curve around the "liberal-democratic trap" is the region corresponding to the shift from the mutual opposition (O,O) payoff in models 2 to the prisoner's dilemma in model 3, figure 5: workers effectively oppose capitalists' interests and capitalists effectively oppose workers' interests.

4. There is one region of the curve where the functional relation between workers' power and class interests has the same general shape for both workers and capitalists: the upward-sloping section to the right of the liberal-democratic trough. This is the region of maximally stable positive class compromise.
5. As working-class power extends beyond corporatist associative practices, the immediate realization of working-class interests again decline. This region of the curve defines the "transition trough" between capitalism and socialism discussed by Adam Przeworski (1985). Capitalists respond to the threat of losing control over the allocation of capital by disinvesting, shifting investments to other places, or by more organized forms of a "capital strike." This has the effect of provoking an economic decline, which hurts workers' material interests. It is only when workers' associational power increases to the point at which investments can be democratically allocated (in the sense of democratically imposed direction on allocation) that the working-class interest curve once again turns upward. Once there is a full realization of hypothetical democratic socialism, the interests of workers and capitalists are once again maximally divergent.

One More Complexity: Zones of Unattainability

In the practical world of real capitalist societies, not all values within this theoretically defined range are historically accessible. There are two different kinds of exclusion mechanisms that have the effect of narrowing the range of real possibilities. These can be termed *systemic* exclusions and *institutional* exclusions.

Systemic exclusions define parts of the curve that are outside the limits of possibility because of the fundamental structural features of the social system. Specifically, the presence of a *constitutionally secure democracy* removes the fully repressed and atomized working-class part of the curve from the historical stage, and the presence of *legally secure capitalist property rights* removes the democratic socialism part of the curve. This does not mean that there are no historical circumstances in which these zones of the curve might become strategically accessible, but to get there would require a fundamental transformation of the underlying social structural principles of the society.

Institutional exclusions refer to various kinds of historically variable institutional arrangements, formed within the limits determined by the systemic exclusions, which make it difficult or impossible to move to specific regions of the curve. For example, restrictive labor law can make it difficult to extend working-class associational power toward the corporat-

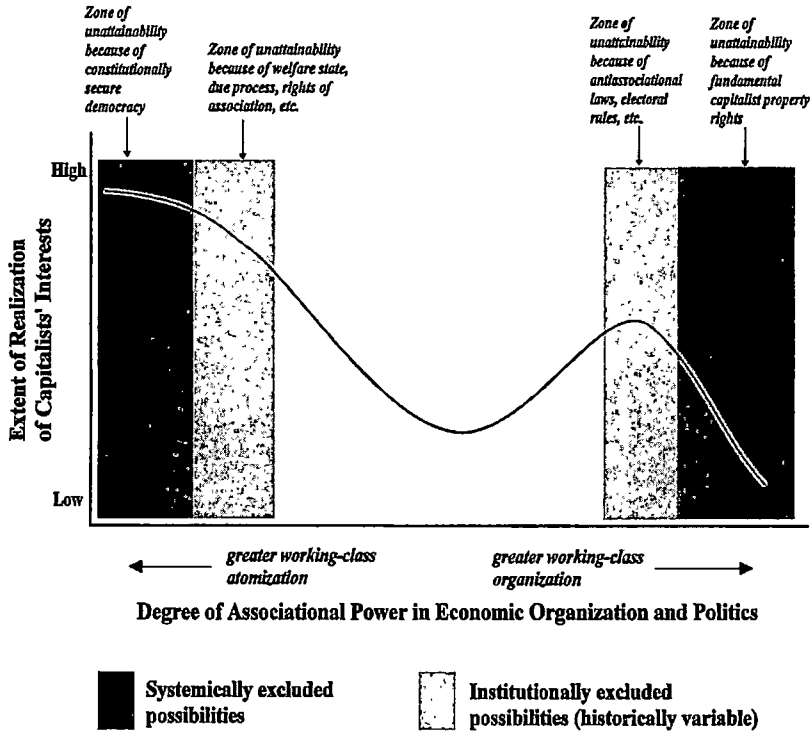


FIG. 12.—Working-class associational power and capitalists' interests in democratic capitalism.

ist associative practices part of the curve (Rogers 1990). On the other hand, generous welfare state provisions, which render workers less dependent on capital, and strong associational rights, which facilitate unionization, may make it difficult to move toward the right-wing managerialist region. Such institutional exclusions, of course, are themselves the outcomes of historical conflicts and should not be viewed as eternally fixed. But once in place, they help to define the range of feasible strategy immediately open to actors, at least until the time when actors can effectively challenge these institutional exclusions themselves.

These two forms of exclusion are illustrated in figure 12. The central region of the curve defines the space that is immediately accessible strategically. To use a game-theory metaphor adopted by Alford and Friedland (1985), this is the domain of ordinary politics, of liberal versus conservative struggles over "plays" within a well-defined set of institutional "rules of the game." The other regions of the curve become the objects of politics only episodically. Reformist versus reactionary politics are struggles over

the rules of the game that define institutional exclusions; revolutionary versus counterrevolutionary politics are struggles over the systemic constraints that define what game is being played.

In figure 12, the zones of unattainability defined by the systemic and institutional exclusions symmetrically span the tails of the theoretical curve of possibilities. There is no reason, of course, to believe that the real world is this neat. Indeed, one of the reasons for introducing this complexity is precisely to provide tools for understanding forms of variation across time and place in these exclusions. This historical variability is illustrated in figure 13, which compares the United States and Sweden in the periods of most stable Swedish social democracy and American liberal democracy.

Systemic exclusions in the United States and Sweden are roughly comparable: both have structurally secure democratic states and capitalist property relations. Where they differ substantially is in the nature of the historically variable institutional exclusions that confront their respective working classes.

In the United States, a variety of institutional rules create a fairly broad band of institutional exclusions to the right of the central trough of the curve. Electoral rules that solidify a two-party system of centrist politics and antiunion rules that create deep impediments to labor organizing all push the boundary of this zone of institutional exclusion to the left (Rogers 1990). On the other hand, such things as the weak welfare state, the very limited job protections afforded workers, and laws that guarantee managerial autonomy all have the effect of narrowing the institutional exclusions centered around right-wing managerialist antiassociational practices. The band of accessible strategy in the United States, therefore, affords very little room to maneuver for labor and keeps working-class associational practices permanently lodged on the downward-sloping segment of the curve to the left of the trough.

Swedish institutional exclusions, particularly during the most stable period of social democracy, all work toward facilitating working-class associational power. Labor law is permissive, making it quite easy to form and expand union membership, and the generous welfare state and job protections significantly reduce the scope of right-wing managerialist strategies. The result has been that the Swedish labor movement has for a long time been located on the upward-sloping section of the curve to the right of the trough.

Actors living within these systems, of course, do not directly see this entire picture. To the extent that the institutional exclusion mechanisms have been securely in place and unchallenged for an extended period of time, they may become entirely invisible, and the parts of the curve that they subsume may become virtually unimaginable. From the vantage

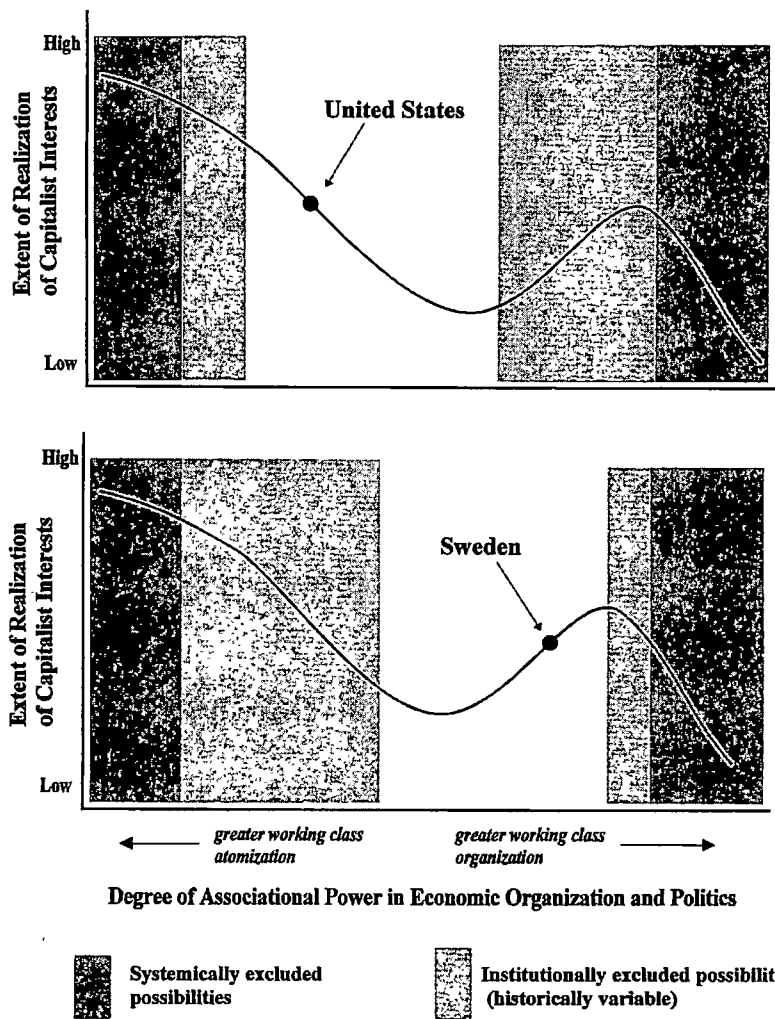
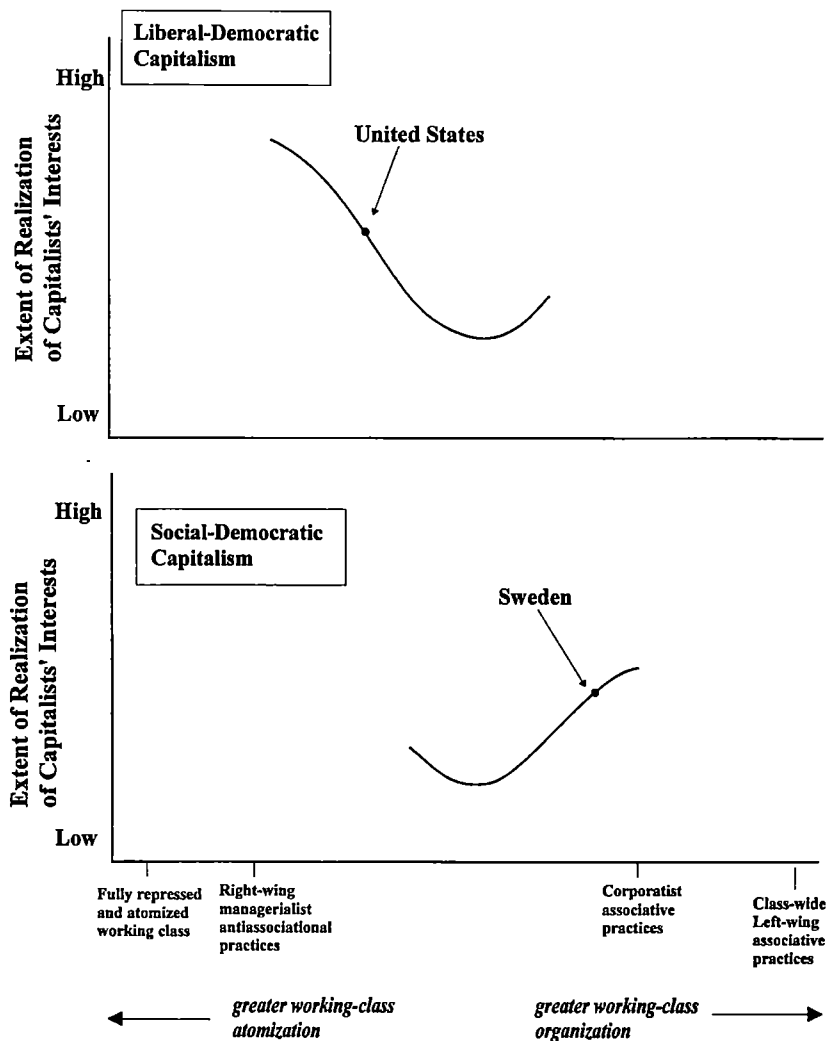


FIG. 13.—Working-class associational power and capitalists' interests in liberal-democratic capitalism (United States) and social-democratic capitalism (Sweden).

point of actors within the system, therefore, the range of “realistic” possibilities may look like those portrayed in figure 14 rather than figure 13. The American labor movement faces a terrain of possibilities that places it chronically on the defensive. Every marginal increase of workers’ strength is experienced by capitalists as against their interests, so when-



Degree of Associational Power in Economic Organization and Politics

FIG. 14.—Strategic environment for feasible associational politics as seen by the actors in liberal-democratic capitalism and social-democratic capitalism.

ever the opportunity arises, capitalists attempt to undermine labor's strength. Antiunion campaigns are common, and decertification elections are a regular occurrence. In Sweden, at least until recently, the institutionally delimited strategic environment is much more benign for workers. The central pressure on capitalists has been to forge ways of effectively cooperating with organized labor, of creating institutional spaces in which the entrenched forms of associational power of workers can be harnessed for enhanced productivity. This need not imply that employers actively encourage working-class associational power, but it does suggest less sustained effort to undermine it.

The immediately accessible strategic environments of workers' struggles for associational power, as illustrated in figure 14, should not be viewed as fixed by an unalterable historical trajectory. The range of attainable possibilities can change, both as the result of conscious political projects to change institutional exclusions and as the result of dynamic social and economic forces working "behind the backs" of actors. Institutional exclusions are created by victories and defeats in historically specific struggles; they can potentially be changed in a similar fashion. But equally, dynamic changes within economic structures can potentially change the shape of the curve itself. It is to that issue that I now turn in a more speculative manner.

TRANSFORMATIONS OF THE TERRAIN OF CLASS COMPROMISE

If the general model of class compromise we have been exploring is reasonably on target, then this suggests that the prospects of positive class compromise can be altered through three different routes: (1) by changes in the *basic shape of the curve*; (2) by changes in the *institutional rules of the game*, which determine the range of institutional exclusions; and (3) by changes in the *associational power of workers* within the strategically accessible range of possibilities.

The first of these involves the functional relations depicted in figures 7 and 8, the second involves the institutional zones of unattainability in figure 12, and the third concerns the specific location within a strategic space as in figure 14. A full-blown theory of class compromise, then, would provide an account of the causal processes that generate these three kinds of effects.

I cannot offer such an elaborated theory. What I will do is propose some relatively speculative hypotheses about the way certain developments in contemporary capitalism may be affecting the first of these elements in the model of class compromise, the overall shape of the reverse-J curve at the heart of the model. Specifically, I will focus on the possible impact of increasing international competition and globalization of capital on the shape of the curve.

For purposes of understanding the changing conditions for positive class compromise, the critical part of the curve in figure 7 is the upward-sloping segment in which working-class associational power positively helps capitalists solve various kinds of collective action and coordination problems. Figure 15 presents a set of tentative hypotheses about how globalization and increasing international competition might affect the relationship between workers' power and capitalists' interests within each of the three institutional spheres of class compromise.

Consider first the effects of globalization on the sphere of exchange. One of the standard arguments in discussions of globalization is that the increasing mobility of financial capital and globalization of markets has undermined the "Keynesian" solutions to macroeconomic problems in advanced capitalism. To the extent that the market for the commodities of capitalist firms are increasingly global, the realization of the economic value of those commodities depends less upon the purchasing power of workers in the countries within which those firms are located. Furthermore, heightened international competition and the constant threats by employers to move production abroad has served to reduce wage pressures, thus reducing the positive effect of strong unions on the problem of wage restraint. The positive effects on capitalists' interests of strong, centralized labor unions has thus probably been reduced by globalization. While there may still be some positive value for capitalists of a strong labor movement in terms of collective action problems of predictable, well-ordered labor markets, especially with respect to the problem of skill formation, nevertheless it seems that, on balance, globalization is likely to depress the positive slope of the first curve of figure 15.

Globalization may have quite different effects in the sphere of production. The characteristic form of working-class associational power within production are works councils and other forms of organized workers' representation within the process of production. As already noted, strong works councils may serve employer interests in a variety of ways: they may increase productivity through greater worker loyalty; they may help spot problems and improve quality control; they may increase the willingness of workers to accept flexible job classifications and work assignments; and they may facilitate the process of intrafirm skill formation. Under conditions of the intensified competition that comes from increased globalization, the positive impact of each of these effects could increase. *If*, therefore, there are significant untapped sources of increased productivity obtainable through enhanced cooperation at the point of production, and *if* working-class associational power within production facilitates such cooperation, then the upward-sloping part of the class compromise curve

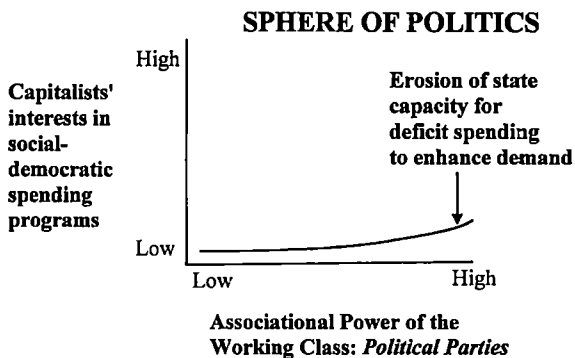
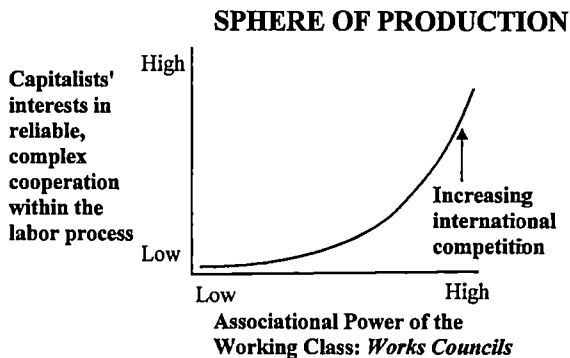
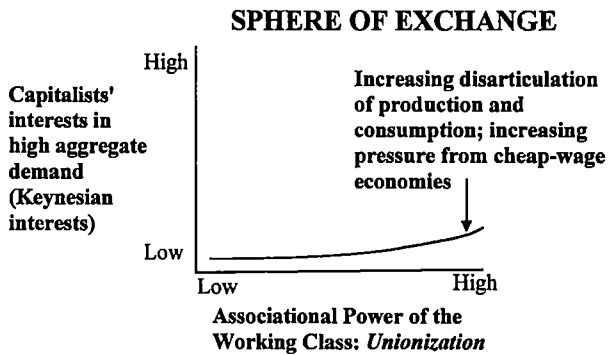


FIG. 15.—Hypothesized effects of globalization and increasing competition of the class compromise curves in the spheres of exchange, production, and politics.

within the sphere of production may rise more steeply as a result of increased competitive pressures.²²

The extent to which these effects on the prospects of positive class compromise in the sphere of production occur depends upon the nature of technology, the labor process, and the organization of work. As Streeck (1991, 1992), Aoki (1988), and others have argued, the maximally productive use of advanced technologies often requires higher levels of information coordination, problem solving, and adaptability than in traditional mass production. To the extent that strong working-class associational power within the sphere of production enhances the levels of trust between employees and managers, and such trust is necessary for such new forms of work organization, the positive effect of workers' power on capitalists' interests may be strengthened. On the other hand, if the technological conditions of production foster weak interdependencies among workers within highly atomized labor processes, increased globalization and competitive pressures would probably not enhance the positive effects of workers' associations within production. This suggests that there are probably strong interactive effects (rather than merely additive effects) of globalization and technological change on the conditions for class compromise within production.

Perhaps the most commonly told story about the negative effects of globalization on the prospects of positive class compromise concerns the sphere of politics. Because of the heightened international mobility of capital, especially financial capital, the argument goes, the capacity of states to engage in deficit spending and other reflationary policies has eroded (Stewart 1984). One of the key ways in which a politically well-organized working class positively benefited capitalists in the past was by creating the conditions for expansive state spending programs that bolstered aggregate demand. The reduced fiscal autonomy of the state resulting from increased globalization both reduces the benefits from such policies and the capacity of the state to sustain them, and thus reduces the positive slope of the class compromise in the sphere of politics.

If these arguments are roughly correct, then this suggests that there will be a tendency under conditions of globalization for positive class compro-

²² Again, just to reiterate the central argument behind the reverse-J curve: works councils, like all forms of working-class associational power, also have negative effects on capitalists' interests. Works councils impose various kinds of rigidities on employers, which interfere with their capacity to unilaterally reorder production in the face of competitive pressures. The downward-sloping curve—not shown in fig. 15—might therefore descend more precipitously under conditions of intensified global competition. The claim here, then, is *not* that the *net* effect of globalization necessarily enhances the value of institutions like works councils, but simply that the positive effects become stronger.

mises, if they are to occur at all, to be more heavily concentrated within the sphere of production in contrast to earlier periods in which positive class compromises were particularly institutionalized within the spheres of exchange and politics. This in turn is likely to generate tendencies toward an intensification of dualism within developed capitalist economies in which some sectors of the labor force are in a position to forge productivist class compromises while others are not.

Dualistic tendencies, of course, are not a new phenomenon. In the 1960s and 1970s there was much discussion of dual labor markets and the division between the "monopoly" and "competitive" sectors of the economy. But, in the past, class compromises within the spheres of exchange and politics sometimes had the effect of muting the effects of such dualisms. Strong unions helped to create wage norms that diffused throughout the economy, benefiting workers not in the most organized sectors, and social-democratic class compromises in the state underwrote a social wage that partially decommodified labor power, again benefiting all workers. The erosion of conditions for stable class compromise in the sphere of exchange and in the state, therefore, risks eroding these countervailing forces to deepening dualism.

CONCLUSION

This article has attempted to chart out a general, abstract model of class compromise in capitalist society. The core intuition builds on Gramsci's insight that in democratic capitalist societies the capitalist class is often hegemonic, not merely dominant, and this implies that class conflict is contained through real compromises involving real concessions, rather than brute force. The bottom line of the argument is that the stability and desirability of such compromises depends upon the specific configurations of power and interests that characterize the relationship between the capitalist class and the working class: when it is the case that working-class associational power positively contributes to solving problems faced by capitalists, then such compromises will be much more durable than when they emerge simply from capacity of workers to impose costs on capitalists.

The theory of class compromise proposed here is thus in keeping with the traditional core of Marxist theory in arguing that the power and struggle are fundamental determinants of distributional outcomes in capitalist societies. But contrary to traditional Marxist ideas on the subject, the model also argues that the configuration of capitalist and worker interests within the "game of class struggle" is not simply determined by capitalism itself, but depends upon a wide variety of economic, institutional, and political factors. Above all, the model argues that class power not only

can affect the outcome of class conflict, but the nature of the game itself: whether or not the confrontation of capital and labor takes the form of a sharply polarized zero-sum conflict or an iterated prisoner's dilemma, or perhaps, a strategic context with significant assurance game features conducive to positive class compromise.

REFERENCES

- Akerlof, George. 1982. "Labor Contracts as Partial Gift Exchange." *Quarterly Journal of Economics* 97 (2): 543–69.
- Akerlof, George, and Janet Yellen. 1986. *Efficiency Wage Models of the Labor Market*. New York: Cambridge University Press.
- Alchian, Armen, and Harold Demsetz. 1972. "Production, Information Costs and Economic Organization." *American Economic Review* 52:777–95.
- Alford, Robert, and Roger Friedland. 1985. *The Powers of Theory: Capitalism, State and Democracy*. Cambridge: Cambridge University Press.
- Aoki, Masahiko. 1988. *Information, Incentives and Bargaining in the Japanese Firm*. Cambridge: Cambridge University Press.
- Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.
- Bartlett, Randall. 1989. *Economics and Power*. Cambridge: Cambridge University Press.
- Bowles, Samuel. 1985. "The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian and Marxian Models." *American Economic Review* 75:16–36.
- Bowles, Samuel, and Herbert Gintis. 1990. "Contested Exchange." *Politics and Society* 18 (2): 165–222.
- . 1998. *Recasting Egalitarianism: New Rules for Communities, States and Markets*. Vol. 3 of *Real Utopias Project*, edited by Erik Olin Wright. London: Verso.
- Calmfors, Lars. 1993. "Centralization of Wage Bargaining and Macroeconomic Performance—A Survey." *OECD Economic Studies* 21 (winter): 161–91.
- Calmfors, L., and J. Driffill. 1988. "Bargaining Structure, Corporatism and Macroeconomic Performance." *Economic Policy* 6:13–61.
- Commons, John R. (1950) 1970. *The Economics of Collective Action*. Madison: University of Wisconsin Press.
- Crouch, Colin, and Wolfgang Streeck, eds. 1997. *Political Economy of Modern Capitalism: Mapping Convergence and Diversity*. Thousand Oaks, Calif.: Sage.
- Doeringer, Peter, and Michael Piore. 1985. *Internal Labor Markets and Manpower Analysis*, 2d ed. Lexington Mass.: D.C. Heath & Co.
- Dore, Ronald. 1973. *British Factory, Japanese Factory: The Origins of National Diversity in Industrial Relations*. Berkeley: University of California Press.
- Edwards, Richard. 1979. *Contested Terrain: The Transformation of the Workplace in the Twentieth Century*. New York: Basic Books.
- Esping-Andersen, Gøsta. 1990. *Three Worlds of Welfare Capitalism*. Princeton, N.J.: Princeton University Press.
- Foulkes, F. K. 1980. *Personnel Policies in Large Nonunion Companies*. Englewood Cliffs, N.J.: Prentice Hall.
- Freeman, Richard. 1988. "Labour Market Institutions and Economic Performance." *Economic Policy* 6:64–80.
- Garrett, Geoffrey. 1998. *Partisan Politics in the Global Economy*. Cambridge: Cambridge University Press.

- Glynn, Andrew. 1995. "Social Democracy and Full Employment." *New Left Review* 211:33–55.
- Goldthorpe, John. 1982. "On the Service Class." Pp. 162–85 in *Social Class and the Division of Labor*, edited by Anthony Giddens and Gavin Mackenzie. Cambridge: Cambridge University Press.
- Gordon, David M. 1976. "Capitalist Efficiency and Socialist Efficiency." *Monthly Review* 29:19–39.
- . 1996. *Fat and Mean*. New York: Free Press.
- Gramsci, Antonio. 1971. *Selections from the Prison Notebooks*, edited and translated by Quintin Hoare and Geoffrey Nowell Smith. New York: International Publishers.
- Greenwald, Bruce. 1979. *Adverse Selection in the Labor Market*. New York: Garland Press.
- Hicks, Alexander, and Lan Kenworthy. 1998. "Cooperation and Political Economic Performance in Affluent Democratic Capitalism." *American Journal of Sociology* 103 (6): 1631–72.
- Kenworthy, Lane. 1995. *In Search of National Economic Success: Balancing Competition and Cooperation*. Thousand Oaks, Calif.: Sage.
- Korpi, Walter. 1983. *The Democratic Class Struggle*. London: Routledge & Kegan Paul.
- Kotz, David. 1994. "Interpreting the Social Structure of Accumulation Theory." Pp. 51–70 in *Social Structures of Accumulation*, edited by David Kotz, Terrance McDonough, and Michael Reich. Cambridge: Cambridge University Press.
- Nakane, Chie. 1970. *Japanese Society*. London: Weidenfeld & Nicholson.
- Olson, Mancur. 1982. *The Rise and Decline of Nations*. New Haven, Conn.: Yale University Press.
- Ouchi, W. G. 1981. *Theory Z: How American Business Can Meet the Japanese Challenge*. Reading, Mass.: Addison-Wesley.
- Pohjola, Matti. 1992. "Corporatism and Wage Bargaining." Pp. 44–81 in *Social Corporatism: A Superior Economic System?* edited by Jukka Pekkarinen, Matti Pohjola, and Bob Rowthorn. Oxford: Clendon Press.
- Pontusson, Jonas. 1997. "Between Neo-Liberalism and the German Model: Swedish Capitalism in Transition." Pp. 50–70 in *Political Economy of Modern Capitalism: Mapping Convergence and Diversity*, edited by Colin Crouch and Wolfgang Streeck. Thousand Oaks, Calif.: Sage.
- Przeworski, Adam. 1985. *Capitalism and Social Democracy*. Cambridge: Cambridge University Press.
- Roemer, John. 1994. *A Future for Socialism*. Cambridge, Mass.: Harvard University Press.
- . 1996. *Equal Shares: Making Market Socialism Work*. Vol. 2 of *The Real Utopias Project*, edited by Erik Olin Wright. London: Verso.
- Rogers, Joel. 1990. "Divide and Conquer: Further 'Reflections on the Distinctive Character of American Labor Law'." *Wisconsin Law Review* 13:1–147.
- Rogers, Joel, and Wolfgang Streeck. 1994. "Productive Solidarities: Economic Strategy and Left Politics." Pp. 128–45 in *Reinventing the Left*, edited by David Miliband. Cambridge: Polity Press.
- Rowthorn, Bob. 1992. "Corporatism and Labour Market Performance." Pp. 82–131 in *Social Corporatism: A Superior Economic System?* edited by Jukka Pekkarinen, Matti Pohjola, and Bob Rowthorn. Oxford: Clendon Press.
- Schmitter, Philippe. 1988. "Corporatism is Dead! Long Live Corporatism! Reflections on Andrew Schonfield's *Modern Capitalism*." *Government and Opposition* 24:54–73.
- Schmitter, Philippe, and G. Lembruch, eds. 1979. *Trends towards Corporatist Intermediation*. London: Sage.

- Shapiro, Carl, and Joseph Stiglitz. 1984. "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review* 74:203-24.
- Sorensen, Aage. 1994. "Forms, Wages and Incentives." Pp. 504-28 in *Handbook of Economic Sociology*, edited by Neil Smelser and Richard Swedberg. Princeton, N.J.: Princeton University Press.
- Soskice, David. 1990. "Wage Determination: The Changing Role of Institutions in Advanced Industrialized Countries." *Oxford Review of Economic Policy* 6:36-61.
- Stewart, Michael. 1984. *The Age of Interdependence*. Cambridge, Mass.: MIT Press.
- Streeck, Wolfgang. 1991. "On the Institutional Conditions for Diversified Quality Production." Pp. 21-61 in *Beyond Keynesianism: The Socio-Economics of Production and Employment*, edited by E. Matzner and W. Streeck. London: Edward Elgar.
- . 1992. *Social Institutions and Economic Performance: Studies of Industrial Relations in Advanced Capitalist Economies*. Newbury Park, Calif.: Sage.
- Streeck, Wolfgang, and Philippe Schmitter. 1985. "Community, Market, State—and Associations? The Prospective Contribution of Interest Governance to Social Order." *European Sociological Review* 1:119-38.
- Waldman, Michael. "Job Assignments, Signaling and Efficiency." *Rand Journal of Economics* 15:255-67.
- Williamson, Oliver. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.
- Wright, Erik Olin. 1985. *Classes*. London: Verso.
- . 1997. *Class Counts: Comparative Studies in Class Analysis*. Cambridge: Cambridge University Press.

Resources and Resourcefulness: Strategic Capacity in the Unionization of California Agriculture, 1959–1966¹

Marshall Ganz
Harvard University

Why did the insurgent United Farm Workers (UFW) succeed while its better-resourced rival—the Agricultural Workers Organizing Committee, AFL-CIO (AWOC)—failed? Explanations relying on altered political opportunity structures or resources, accounts of Cesar Chavez’s charismatic leadership, or descriptions of UFW strategy fail to identify mechanisms for creating effective strategy. By analyzing leadership, organizational influences on actors’ choices, and their interaction within the environment, this study shows that greater access to salient information, heuristic facility, and motivation generated more effective strategy. Differences in “strategic capacity” can explain how resourcefulness can compensate for lack of resources, why some new organizations can overcome the “liability of newness,” and how reorganizational “focal” moments may lead to a social movement.

The sociological imagination enables us to grasp history and biography and the relations between the two within society.

(C. Wright Mills, *The Sociological Imagination*)

INTRODUCTION

Until the success of the United Farm Workers (UFW) in the 1960s and 1970s, repeated efforts to unionize California’s 400,000 farmworkers had failed.² Since 1900, three major attempts at labor organization had been

¹ I wish to thank Theda Skocpol, Richard Hackman, Susan Eaton, Elisabeth Clemens, John Campbell, William Gamson, Mark Moore, Doug McAdam, Jorge Dominguez, David Hart, Paul Osterman, Robert Putnam, Kenneth Andrews, Ziad Munson, Andy Molinsky, Mark Warren, and the *AJS* reviewers for their thoughtful comments on earlier versions of this article. Direct correspondence to Marshall Ganz, Department of Sociology, Harvard University, 33 Kirkland, William James Hall 473, Cambridge, Massachusetts 02138. E-mail: ganz@wjh.harvard.edu

² The union which has called itself the United Farm Workers (UFW) since 1972 called itself the Farm Workers Association (FWA) from 1962 to 1964, the National Farm Workers Association from 1964 to 1966 (NFWA), and the United Farm Workers Organizing Committee (UFWOC) from 1966 to 1972. When referring to the organiza-

made by four different unions: the Industrial Workers of the World (IWW) prior to World War I; the Cannery, Agricultural, and Industrial Workers Union (CAIWU) and United Cannery, Agricultural Packing, and Allied Workers CIO (UCAPAWA) in the 1930s; and the National Farm Labor Union AFL (NFLU) in the 1940s. By the spring of 1966, however, just four years after it had begun organizing and six months after calling its first strike, the UFW signed the first multiyear union contract to cover California farmworkers, established a new union, and launched a farm-worker movement. By 1977, the UFW held over 100 contracts, had recruited a dues-paying membership of more than 50,000, and had secured enactment of the 1975 California Agricultural Labor Relations Act, the only legislative guarantee of collective bargaining rights for agricultural workers in the continental United States. The UFW had also played a major role in the emergence of a Chicano movement in the Southwest, had recruited and trained hundreds of community activists, and had become a significant player in California politics. In this article, I ask why the newly formed, uncertainly funded, and independent United Farm Workers succeeded, while the well-established union with which it found itself in competition failed—the AFL-CIO's Agricultural Workers Organizing Committee (AWOC).³

Answering this question requires going beyond Jenkins and Perrow's (1977) classic of social movement theory, "Insurgency of the Powerless," which attributes the UFW's success, as compared with another effort in the late 1940s, to a more favorable political opportunity structure. Although the 1960s were an era of greater political responsiveness to groups like the UFW than the 1940s had been, this cannot explain why the UFW was more successful than its rival. Nor can scholars who rely on differences in resources to account for differences in outcomes (McCarthy and Zald 1977) explain why the resource-poor UFW succeeded while its resource-rich opponents did not. Those who attribute the UFW's success to the "charismatic leadership" of Cesar Chavez (Nelson 1966; Dunne 1967; Mathiessen 1969) explain little of what that is, where it comes from, and why it works. And even those who acknowledge the unique effectiveness of the UFW's strategy (Brown 1972; Majka and Majka 1982; Jenkins 1985) do not explain why it was the UFW that developed this strategy

tion in general, I use the designation "UFW," but when referring to it in a specific historical context, as I do in most of this article, I use the name that was used at the time.

³ This article only deals with the period of competition between the UFW and the AWOC. Competition from the Teamsters began in 1966 and continued, off and on, until they withdrew in 1977. Elsewhere, I show that similar differences in strategic capacity account for UFW success vis-à-vis the Teamsters.

and not its competitor. Drawing evidence from this case and insights from organization theory, social psychology, and cognitive sociology, I argue that consigning the influence of leadership and organization on strategy to a “black box” has created a serious deficit in social movement theory. As shown in figure 1, I argue that difference in the outcomes of AWOC and UFW efforts can be explained by differences in their strategy—the targeting, timing, and tactics through which they mobilized and deployed resources. Differences in their strategy, however, and the likelihood it would be effective in achieving desired goals, were due to differences in leaders’ access to salient information about the environment, heuristic use they made of this information, and their motivation—what I call their “strategic capacity.” Differences in strategic capacity, in turn, were due to differences in leaders’ life experience, networks, and repertoires of collective action and the deliberative processes, resource flows, and accountability structures of their organizations. Strategic capacity is greater if a leadership team includes insiders and outsiders, strong and weak network ties, and access to diverse, yet salient, repertoires of collective action and also if an organization conducts regular, open, authoritative deliberation, draws resources from multiple constituencies, and roots accountability in those constituencies. I explain the UFW’s success over its rivals by differences in its strategy, account for differences in strategy by the way in which it was developed, and explain how it was developed in the interaction of leadership and organization with environment.

Analysis of this case offers a way to specify conditions under which one organization is more likely than another to develop strategy that is effective in achieving its goals. I do not focus on why one kind of tactic is more effective than another—a topic dealt with extensively in political, military, and management literature—but rather on why one organization is more likely to develop a series of effective tactics than another—its strategic capacity. Although strategic capacity, strategy, and outcomes are distinct links in a probabilistic causal chain, I argue greater capacity is likely to yield better strategy, and better strategy is likely to yield better outcomes. Differences in strategic capacity can explain how resourcefulness compensates for lack of resources when insurgent social movements overcome more powerful opponents. Differences in strategic capacity can also explain why some new organizations overcome the “liability of newness” (Stinchcombe 1965) to succeed in domains in which old organizations fail, suffering from a “liability of senescence.” Because of more recent selection, organizational flexibility, and closer articulation with the environment, leaders of new organizations may have greater access to salient information about current environments, heuristic opportunity, and motivation than leaders of old organizations. Attention to strategic capacity also helps explain why reconfigurations of leadership and organization

at certain “focal moments” (Morris 1993; Sewell 1996; Lofland 1996) can produce dramatic strategic change, including the emergence of a social movement.

METHODOLOGY

This case study compares a sequence of concurrent strategic choices made by the AWOC and the UFW at three critical junctures—their organizational foundings, the Delano Grape Strike, and the Schenley Boycott. Because it provides an opportunity to “control” for environment, this case allows us to bring full attention to differences in the actors and how they interacted with their environment. A case study with multiple points of comparison can be an especially useful way to discern the underlying mechanisms that may account for repeated differences in outcome (Campbell 1975).

Although differences in the outcomes of these efforts are obvious, the factors that influenced the outcomes must be independently observable to have explanatory value and a mechanism shown through which they operate (King, Keohane, and Verba 1994). Drawing on archival data, interviews, and participant observation, I examine the differences in each organization’s strategy and its influence on outcomes by comparing differences in targeting, timing, and tactics at three critical moments—at each of which the UFW achieved its goals more effectively than did the AWOC.

Having established a link between strategy and outcomes, I draw on organization theory, social psychology, and cognitive sociology—which have contributed to social movement theory in the past (McCarthy and Zald 1977; Gamson, Fireman, and Rytina 1982; Snow, Rochford, Worden, and Benford 1986)—to conduct a grounded analysis of my data to identify conditions that could account for such dramatic differences in strategy. Because strategy unfolds in the interaction of leaders, organizations, and environment, I focus on sites of this interaction. I compare leaders as to their biographies, sociocultural networks, and tactical repertoires, and organizations as to their deliberative arrangements, resource flows, and accountability structures. This comparison reveals starkly different patterns of leadership and organization with direct bearing on formulation of strategy. Probing to discern the mechanism by which these differences turned into strategy drew my attention to patterns in the kind of information to which leaders had access, their heuristic use of this information, and their motivation, which I now call strategic capacity.

The data on which this analysis is based is drawn from primary and secondary sources as well as my experience with the UFW from 1965 to 1981 as an organizer, organizing director, and national officer. This raises

a potential problem of bias based on my personal experiences, interests I may have in particular accounts of controversial events, and personal relationships with persons on all sides of the conflict. But my experience also equips me with a deep understanding of the context of these events, direct information as to what took place, and access to important research resources. In an attempt to realize the benefits while minimizing the risks, I “triangulate” my data for this study by drawing on multiple primary and secondary sources, relying on my own experience only where specifically noted. Primary sources include interviews with participants; newspaper accounts, including those published by labor journals; published memoirs of participants; archival materials, including published proceedings of meetings; and my own notes and personal records. Secondary sources include histories of farm labor organizing; scholarly and journalistic accounts of the activities of the AWOC, the UFW, and the Teamsters; and unpublished doctoral dissertations.

Rather than attempting to test the influence of a single variable, the purpose of this project is to generate a grounded, yet theoretically informed, hypothesis to account for the observed differences in outcomes—and which can be further tested in other settings (Glaser and Strauss 1967; Liphart 1971; Skocpol 1984; Little 1991; Parry 1998). Because I develop a model to explain outcomes based on the interaction of a complex set of variables, my approach raises a problem of overdetermination or how to distinguish what actually “caused” what. I do not claim to have found a “crucial” variable necessary or sufficient to account for the differences in observed outcomes. Rather, I argue the outcome I explain—one group repeatedly develops more effective strategy than another—is more or less likely to the extent conditions specified in this model are met. In poker, chance may determine the outcome of any one hand, or even a game, but, in the long run, some players are more likely to be winners than others. An organization can always stumble on an opportunity, but I argue the likelihood it will make strategic use of it depends on factors I specify here. In this way, I explain the UFW’s success compared with its rival by the greater effectiveness of its strategy, account for difference in the effectiveness of its strategy by the way it was developed, and explain the way it was developed by the interaction of leadership and organization with environment.

THEORY

Strategy and Social Movements

Social movement scholars tell us little about the relationship of strategic leadership to social movement success. Explanations of the emergence, development, and outcomes of social movements are usually based on

variation in resources, opportunities, and framing (McAdam, McCarthy, and Zald 1996)—concepts that stress the influence of environmental change on actors. In this view, social movements unfold as actors predictably respond to new political opportunities, newly available resources, or changes in cultural frames. Although students of tactics do offer accounts of their sources (Tilly 1981; Freeman 1979) and their effect on outcomes (Lipsky 1968; Gamson 1975; McAdam 1983), they do not explain why one organization should devise tactics that turn out to be more or less strategically effective than those of another. Even in domains of management, military, and political studies in which strategic leadership receives far greater attention, the focus has been more on how leadership and strategy work than on explaining why some organizations devise more effective strategy than others.

Understanding social movements, however, requires accounting for the fact that different actors act in different ways, some of which influence the environment more than others. Some see political opportunities where others do not, mobilize resources in ways others do not, and frame their causes in ways others do not. And because it is based on the innovative, often guileful, exercise of agency, good strategy is often anything but obvious. It can thus be hard to deduce from “objective” configurations of resources and opportunities and is based rather on novel assessments of them (Morris 1984; Gamson and Meyer 1996; Jasper 1997). As a consequence, popular accounts of effective strategy attribute it to the charismatic gifts of particular leaders rather than offering systematic explanations of conditions under which leaders are more or less likely to devise effective strategy.

But neither is strategy purely subjective. Strategic thinking is reflexive and imaginative, based on how leaders have learned to reflect on the past, pay attention to the present, and anticipate the future (Bruner 1990). Leaders are influenced by life experience, relationships, and practical learning, which provide them with lenses through which they see the world (Bandura 1989; Powell and DiMaggio 1991; Banaszak 1996; Zerubavel 1997; DiMaggio 1997), and by the organizational structures within which they work and through which they interact with their environment (Weick 1979).⁴

Failure to bring attention to the important influence of strategic leader-

⁴ Sociologists have begun to describe this approach as a “cognitive sociology,” which they distinguish from “cognitive universalism”—in which actors are assumed to react similarly to similar environmental stimuli—and “cognitive individualism”—in which actors are assumed to react entirely idiosyncratically (Zerubavel 1997). This distinction is similar to that which social psychologists make between “mechanistic agency,” “autonomous agency,” and “emergent interactive agency” (Bandura 1989).

ship on social movements is an important theoretical shortcoming in general—and in scholarly efforts to explain the farmworker movement in particular. Students of the farmworker movement who focus on the changing environment or political opportunity structure correctly point to the importance of an emergent liberal coalition (Jenkins and Perrow 1977; Jenkins 1985) and new politics of race (Majka and Majka 1982) but do not explain why the UFW took advantage of these opportunities in ways its competitors did not. Scholars who rely on differences in resources to account for success and failure (Jenkins and Perrow 1977; Jenkins 1985) cannot explain why the resource-rich AWOC failed, while the resource-poor UFW succeeded. Scholars who point to the UFW's strategy of redefining the arena of conflict by mobilizing dual constituencies of farmworkers and supporters, so they could "turn the moral tables" on the opposition with a grape boycott (Brown 1972; Majka and Majka 1982; Jenkins 1985), fail to explain why only the UFW devised this strategy. And observers who attribute the UFW's success to the charismatic leadership of Cesar Chavez (Nelson 1966; Dunne 1967; Mathiessen 1969) fail to explain what it is, where it came from, and how it worked.

Understanding Strategy

Strategy is the conceptual link we make between the places, the times and ways we mobilize and deploy our resources, and the goals we hope to achieve (Clausewitz 1832; Hamel and Prahalad 1989; Porter 1996; Eisenhardt and Brown 1998). Strategy is how we turn what we have into what we need—by translating our resources into the power to achieve purpose. Although we often do not act "rationally" and outcomes are often unintended (Cohen, March, and Olsen 1972; Salancik and Pfeffer 1977), we do act purposefully (Weick 1979; Watson 1990). Strategy is effective when we realize our goals by means of it. Studying strategy is a way of discerning pattern in the relationship between intention, action, and outcome.

Strategy is a way of "framing" specific choices about targeting, timing, and tactics. As schema theory teaches us, we attribute meaning to specific events by locating them within broader frameworks of understanding (Goffman 1974; Snow et al. 1986; Fiske and Taylor 1991; Gamson 1992; D'Andrade 1992; Gamson and Meyer 1996; DiMaggio 1997). The strategic significance of choices we make about how to target our resources, time our initiatives, and employ tactics depends on how we frame them in relation to other choices in a path toward our goals (Watson 1990). One reason it is difficult to study strategy is that, although choices about targeting, timing, and tactics can be directly observed, the strategic "frame" within which we make these choices—and provide them with their coherence—must often be inferred.

Since strategy is a way of orienting *current* action toward future goals, however, it develops in interaction with an ever-changing environment, especially the actions and reactions of other actors (Alinsky 1971; Weick 1979; J. B. Quinn 1980; Mintzberg 1987; Burgelman 1991; Mintzberg 1994; Hamel 1996; Brown and Eisenhardt 1997). In fixed contexts in which rules, resources, and interests are given, strategy can to some extent be understood in the analytic terms of game theory (Schelling 1960). But in settings in which rules, resources, and interests are emergent—such as social movements—strategy has far more in common with creative thinking (Morris 1984; Hamel 1996; Brown and Eisenhardt 1998). Strategic thinking in these settings can best be understood as an ongoing creative or innovative process of understanding and adapting new conditions to one's goals (Brown and Eisenhardt 1998).

My argument about the relationship of strategy to outcomes can be clarified by drawing on the distinction game theorists make between games of chance, skill, and strategy (Schelling 1960). In games of chance, winning depends on the luck of the draw. In games of skill, it depends on behavioral facility, like hitting a tennis ball. In games of strategy, it depends on cognitive discernment—in interaction with other players—of the best course of action, as in Go. In most games, all three elements come into play. In poker, for example, getting cards is a matter of chance; estimating probabilities, of skill; and betting, of strategy. Although chance may be dispositive in any one hand, or even one game, in the long run, skill and strategy distinguish excellent players—and their winnings—from others. Similarly, environmental developments can be seen as “chance,” in so far as any one actor is concerned. But, in the long run, some actors are more likely to achieve their goals than others because they are better able to take advantage of these chances. Environmental changes may generate opportunities for social movements to emerge, but the outcomes and legacies of such movements have far more to do with strategies actors devise to turn these opportunities to their purposes—thus reshaping their environment.

Strategic Capacity

Organizations differ in the likelihood they will develop effective strategy—what I call their “strategic capacity.” Viewing strategy as a kind of creative or innovative thinking, I build on the work of social psychologists, cognitive sociologists, and organization theorists by focusing on three key influences on creative output: salient knowledge, heuristic processes, and motivation (Amabile 1996).⁵ In this section, I link leadership and organiza-

⁵ I am particularly indebted to Amabile's fine work on creativity, which provided me with an important link between the microbehaviors and macro-outcomes I am trying

tional variables with these elements to specify conditions under which one group is likely to develop more effective strategy than another (see table 1, below).

Salient knowledge.—When actors face routine problems, their familiarity with domain-specific algorithms—or action repertoires—facilitates effective problem solving. The better one's information about a domain within which one is working, the better the "local" knowledge, the more likely one is to know how to deal effectively with problems that arise within that domain. Since environments change in response to actors' initiatives, however, regular feedback is especially important in evaluating responses to these initiatives (Zaltman, Duncan, and Holbeck 1973). The test of salience I use is one of relevance to the environment in which leaders are trying to get results, or their "operating environment."

Heuristic processes.—When faced with novel problems—often the case for leaders of organizations operating within new or changing environments—heuristic processes permit actors to use salient knowledge to devise novel solutions by imaginatively recontextualizing their understanding of the data.⁶ This reframes understanding of the data so as to make alternative interpretations and pathways conceivable, facilitating analogic thinking (Lakoff and Johnson 1980; Gentner 1989; Abelson 1981; White 1992; Strang and Meyer 1994) and bricolage (Lévi-Strauss 1966; Douglas 1986; Campbell 1997). At the most basic level, the more different ideas are generated, the greater the likelihood there will be good ones among them (Campbell 1960; Simonton 1988). Encounters with diverse points of view and ways of doing things thus facilitate innovation (Kasperson 1978), whether based upon one's life experience (Bernstein 1975; Langer 1989; Rosaldo 1989; Piore 1995) or the experience of a group (Weick 1979; Senge 1990; Rogers 1995; DiMaggio 1997). Knowledge of diverse domains not only offers multiple routines from which to choose,

to explain. In adapting her work, I substitute the term "salient knowledge" for "domain specific skills" to better capture the crucial role of environmental information in strategy. I also focus on "recontextualization" as a key heuristic element in strategic thinking and consider a broader range of motivational sources. Although much of the scholarly work on creativity has been done at the individual level, studies of team and organizational innovation suggest enough similarity to use these studies as a starting point for looking at how leadership teams interact adaptively with each other and their environments to formulate strategy (Boone, van Olffen, and van Witteloostuijn 1998; Hutchins 1991).

⁶ These processes also include "breaking set" during problem solving (Newell, Shaw, and Simon 1962), understanding complexity (E. Quinn 1980), keeping response options open (Getzels and Csikszentmihalyi 1976), suspending judgment (Osborn 1963; Stein 1975), using wide categories (Cropley 1990; Bernstein 1975), breaking out of "scripts" (Langer 1978; Langer and Imber 1979), brainstorming (Osborn 1963), and playfulness with ideas (Wickelgren 1979; March and Olsen 1976).

TABLE 1
THREE ELEMENTS OF STRATEGIC CAPACITY

	Motivation	Salient Information	Heuristic Processes
Dimensions of leadership:			
Biography (insiders and outsiders)	Personal, vocational commitment	Diverse local knowledge	Broad contextualization
Networks (strong and weak ties)	Intrinsic rewards Personal commitment Reputation	Diverse local knowledge Feedback	Broad contextualization
Repertoires (diverse salient repertoires)	Competence Feedback	Diverse local knowledge	Sources of bricolage or analogy
Organization:			
Deliberation (regular, open, and authoritative)	Commitment Autonomy	Diverse local knowledge	Heterogeneous perspectives Periodic assessment
Resource flows (multiple constituencies; reliance on people)	Commitment Autonomy Feedback	Feedback	Heterogeneous alternatives
Accountability (salient constituencies; entrepreneurial or democratic)	Commitment Intrinsic rewards Feedback	Diverse local knowledge Feedback	Heuristic skills

NOTE.—The illustrated influences of leadership and organization on motivation, heuristics, and information are meant to be simultaneous, not sequential.

but contributes to “mindfulness” (Langer 1989) that multiple solutions are possible (Bernstein 1975; Langer 1989; Senge 1990; DiMaggio 1997) and that most known solutions are “equivocal” (Weick 1979).

Motivation.—Motivation is critical to creative output because of its effect on the focus actors bring to their work (Ruscio, Whitney, and Amabile 1995), their ability to concentrate for extended periods of time (Prentky 1980), their persistence (Walberg 1971), their willingness to take risks (Glover and Sautter 1977), and their ability to sustain high energy (Bergman 1979). Motivated actors are also more likely to do the work it takes to acquire needed domain-specific knowledge and skills than those who are less motivated (Conti, Amabile, and Pollack 1995). Perhaps the most important source of creative motivation is the “intrinsic reward” it brings to actors who love their work—for whom it is their “vocation”—in contrast with those motivated by “extrinsic rewards,” which can actually inhibit creativity (Amabile 1996). Actors can also override “programmed” modes of thought to think more critically and reflexively (DiMaggio 1997) if they are intensely interested in a problem (Abelson 1981), dissatisfied with the status quo (Bourdieu 1990), or experience a schema failure as a result of sharp breaches in expectations and outcomes (Moscovici 1984; Garfinkel 1967; Swidler 1986). The influence of motivation on outcomes also helps explain the positive effect of affective or normative commitment on workplace performance (Mowday, Porter, and Steers 1982; Meyer and Allen 1997). Finally, one reason successful leadership teams can become more successful over time is that success augments motivation (Deci and Ryan 1980; Chong 1991), not only resources and opportunities. Organizational settings in which people enjoy autonomy, receive positive feedback from peers and superiors, and are part of a team competing with other teams enhance their motivation. It is diminished when they enjoy little autonomy, get no feedback or negative feedback from peers and superiors, and are competitive within a team (Amabile 1988; Hackman 1990).

Sources of Strategic Capacity

As illustrated in table 1 and explained in the text below, I argue that leadership variables of biography, networks, and repertoires, and organizational variables of deliberation, resource flows, and accountability, link in specific ways to each of the above three elements of an organization’s “strategic capacity.”

Leadership.—To identify sources of variation in strategic capacity, I compare the leaders of the AWOC and the UFW as to their biographies, sociocultural networks, and tactical repertoires—who they were, whom they knew, and what they knew. I define leaders as persons authorized to make strategic choices within an organization (Oberschall 1973; Porter

1996). I do not evaluate their qualities of leadership as such but rather their contribution to formulation of strategy. Although researchers have linked leaders' psychological, professional, organizational, and generational backgrounds to strategies they adopt, few have studied the relationship between leaders' backgrounds and the likelihood they will develop *effective* strategy (Mannheim 1952; Kuhn 1962; Oberschall 1973; Chandler 1977; Freeman 1979; Ross 1983; Lofland 1996).

Although strategy is more often described as the work of individual leaders than of formal or informal leadership teams, I argue that strategy is usually a product of the interaction among those persons who share responsibility for its formulation—what I call here a “leadership team.” I recognize that the “person in charge” plays a uniquely important role in formulating strategy, particularly in the formation and maintenance of the leadership team itself (Hackman and Walton 1986). But I argue that strategy is more often the result of interaction among leaders than organizational myths acknowledge. I argue that leadership teams that combine insiders and outsiders, strong and weak ties to constituencies, and diverse yet salient repertoires of collective action have greater capacity to develop effective strategy than those that do not.

Because biographical experience is the primary source of a person's cognitive socialization (Bernstein 1975; DiMaggio 1997; Zerubavel 1997), cultural perspective (Rosaldo 1989; Jasper 1997), and motivating interests (D'Andrade 1992), I first analyze leaders' biographies as to race, class, gender, generation, ethnicity, religious beliefs, family background, education, and professional training.⁷ As shown in the first row of table 1, I argue that leadership teams of “insiders” and “outsiders” combine diversity of salient local knowledge with an opportunity to heuristically recontextualize this knowledge (Bernstein 1975; Weick 1979; Senge 1990; Rogers 1995; Hamel 1996). Persons with “borderland” experience of straddling cultural or institutional worlds may make innovative contributions for the same reasons (Campbell 1960; Kuhn 1962; Rickards and Freedman 1978; Weick 1979; Rosaldo 1989; Piore 1995). Insiders personally committed to constituencies with whom they identify or outsiders normatively committed to a vocation are likely to be more motivated than those whose interest is solely instrumental or professional (Weick 1979; Howell 1990; Meyer and Allen 1997)—and they are more likely to find their work intrinsically rewarding (Amabile 1996).

As shown in the second row of table 1, teams that combine leaders with “strong” and “weak” ties will have greater strategic capacity than those that do not. Sociocultural networks (White 1992; Emirbayer and Goodwin

⁷ One of the few social movement studies linking leadership with strategy is Wickham-Crowley's (1992) excellent comparison of Latin American guerrilla insurrections.

1994) are sources of ideas about what to do and how to do it (Granovetter 1973), mechanisms through which social movements recruit (Stark and Bainbridge 1985; McAdam and Paulsen 1993), sources of social capital (Coleman 1990; Putnam 1993; Chong 1991), and incubators of new collective identities (Taylor and Whittier 1992; Gamson 1991). Leaders with "strong ties" to constituencies are more likely to possess salient information about where to find resources, whom to recruit, what tactics to use, and how to encourage these constituencies to identify with the organization (Morris 1984). On the other hand, leaders with "weak ties" to diverse constituencies are more likely to know how to access the diversity of people, ideas, and routines that facilitate broad alliances (Granovetter 1973). Combinations of strong and weak ties are associated with social movement recruitment because they link access with commitment (Gamson 1990), just as they are associated with innovation because they link information with influence (Rogers 1995). Informal and formal ties are also important means for feedback of salient information, especially on organizational initiatives. Diverse ties, like diverse life experience, facilitate heuristic "recontextualization" of strategic choices. Strong ties strengthen a leaders' motivation to the extent they have personal commitments to those whose lives are influenced by choices they make and from whom they acquire their reputations (Chong 1991).

Finally, as shown in row three, leaders with knowledge of a diversity of salient collective action repertoires are more likely to develop effective strategy than those without such knowledge (Hamel 1996; Moore 1995; Alexander 1998). Knowledge of collective action repertoires is valuable because of their practical (people know what to do), normative (people think they are right), and institutional (they attach to resources) utility in mobilizing people who are familiar with them (Tilly 1986; Clemens 1996). Repertoires known to one's constituency, but not to one's opposition, are particularly useful (Alinsky 1971). Knowledge of multiple repertoires not only widens leaders' range of possible choices, but also affords them the opportunity to adapt to new situations by heuristic processes of bricolage or analogy. The motivation of leaders adept in these repertoires is enhanced by competence they experience in their use and by positive feedback from constituencies who find these repertoires familiar.

Organization.—Turning to the second major set of influences on strategic capacity, I argue that organizational structures that afford leaders venues for regular, open, and authoritative deliberation; draw resources from a diversity of salient constituencies; and hold leaders accountable to those constituencies—and to each other—are more likely to generate effective strategy than those that do not. Organizational structure is created by commitments among founders who enact ways to interact with each other

and with their environment (Weick 1993). It defines patterns of legitimacy (Weber [1948] 1978; Powell and DiMaggio 1991), power (Emerson 1962; Pfeffer and Salancik 1978; Perrow 1986), and deliberation (March and Olsen 1976). Although organizational form is a consequence of founders' strategic choices (Child 1972; Oliver 1988; Eisenhardt and Schoonhoven 1990; Clemens 1996), once established, it has a profound influence on innovativeness (Zaltman et al. 1973; Damanpour 1991) and strategy (Bower 1970).

As shown in row 4, leaders who take part in regular, open, and authoritative deliberation gain access to salient information, participate in a heuristic process by means of which they learn to use this information, and are motivated by commitment to choices they participated in making and upon which they have the autonomy to act (Duncan 1973; Hackman 1990; Amabile 1996). Regular deliberation facilitates initiative by encouraging periodic assessment of the organization's activities (Brown and Eisenhardt 1997, 1998). Deliberation open to heterogeneous points of view enhances strategic capacity because "deviant" perspectives facilitate better decisions (Nemeth and Staw 1989), encourage innovation (McCleod 1992; Weil 1994), and develop group capacity to perform cognitive tasks more creatively and effectively (Hutchins 1991). Authoritative deliberation—in the sense that it results in actionable decisions—motivates actors both to participate in decisions and to implement that which was decided upon (Hackman 1990).⁸

As shown in row 5, organizations that mobilize resources from multiple constituencies enjoy greater strategic capacity than those that do not. First, leaders who must obtain resources from constituents must devise strategy to which constituents will respond (Chandler 1962; Mansbridge 1986). If membership dues are a major source of support, leaders learn to get members to pay dues. However, reliance on resources drawn primarily from outside the operating environment—even when those resources are internal to their organizations—may dampen leaders' motivation to devise effective strategy. As long as they attend to the politics that keep the bills paid, they can keep doing the same thing "wrong." At the same time, leaders who draw resources from multiple constituencies gain strategic flexibility because they enjoy the autonomy of greater room to maneuver (Powell 1988; Alexander 1998). Finally, a decision to rely more

⁸ To the extent strategy is viewed as emergent and improvisational, as in "adhocracies" in which decision makers are implementers, the distinction between decision making and implementation is problematic (Mintzberg and McHugh 1985). Skilled strategists take advantage of the unexpected, turning it to their purposes, even as it alters their plans.

on people than on money encourages growth in strategic capacity when it encourages selection and development of more leaders who know how to strategize. The more strategists, the greater the flexibility with which an organization can pursue its objectives and the scale on which it can do so (Weick 1979). Leaders' choices as to constituencies from whom to mobilize resources thus strongly influence their subsequent strategy (Oliwer and Marwell 1992).

Finally, as shown in row 6, accountability structures affect strategy by establishing routines for leadership selection and defining loci of responsiveness. Leaders accountable to those outside the operating environment may have been selected based on criteria that have little to do with knowledge of—or motivational connection with—constituencies within that environment. Leaders selected bureaucratically are more likely to possess skills and motivations compatible with bureaucratic success than with the creative work innovation requires. Leaders selected democratically are at least likely to have useful knowledge of the constituency that selected them and enough political skills to have been selected if that constituency is within one's operating environment. Entrepreneurial or self-selected leaders—in the sense that the undertaking is their initiative—are more likely to possess skills and intrinsic motivations associated with creative work (Chambers 1964; MacKinnon 1965; Getzels and Csikszentmihalyi 1976). Although democratic and entrepreneurial leadership selections are in tension with one another, either may yield greater strategic capacity than bureaucratic leadership selection.

Liability of Senescence

It is a "given" of organization theory that the failure rate among new organizations is greater than that among old organizations—the "liability of newness" (Stinchcombe 1965; Freeman, Carroll, and Hannan 1983). At the same time, scholars describe processes of organizational inertia (Hannan and Freeman 1977) that inhibit adaptation by old organizations to new environments, opening "niches" within which new organizations emerge—a liability of aging (Aldrich and Auster 1986) or "senescence." Differences in strategic capacity may explain why some new organizations do survive and at the same time account for less adaptive behavior by older organizations. Leaders of new organizations may have more strategic capacity because they were recently selected, have more organizational flexibility, and work in closer articulation with the environment. Leaders of old organizations often were selected in the past, are constrained by institutionalized routines, and may have the resources that allow them to operate in counterproductive insulation from the environment.

Focal Moments

Finally, because of the profound influence of leadership and organization on strategy, the choices leaders make at “focal moments,” which reconfigure leadership and organization themselves, can create dramatically new strategic possibilities, including conditions for the emergence of a social movement (Smelser 1962; Morris 1993; Lofland 1996; Sewell 1996). As I will show below, this is what took place with the UFW.

Summary

Returning to figure 1, this article argues that an organization is more likely to achieve positive outcomes if it develops effective strategy, and it is more likely to develop effective strategy if its leaders can access diverse sources of salient information, employ heuristic processes, and demonstrate deep motivation—their strategic capacity. Variation in strategic capacity, again, derives from differences in leaders’ life experience, networks, and repertoires, and organizations’ deliberative processes, resource flows, and accountability structures.

ORGANIZING CALIFORNIA’S *FACTORIES IN THE FIELD*

California’s uniquely large-scale agricultural industry requires seasonal workers.⁹ It has also tried to protect itself from seasonal demands those workers make when organized.¹⁰ Growers solved this problem historically by employing workers who were politically, economically, and culturally disfranchised. Politically, their status as migrants, immigrants, or undocumented workers meant few farmworkers ever became voters. Economically, farmworkers were excluded from the 1935 National Labor Relations Act (NLRA), which conferred the right to organize on industrial workers. And culturally, the fact that farmworkers were most often recruited from communities of color placed them beyond the racially bounded concerns of many white Americans (Majka and Majka 1982). As a result, local law enforcement usually stood ready to crush strikes through intimidation,

⁹ This brief account is drawn from a number of histories of California agriculture and agricultural workers, especially Jamieson (1975), McWilliams (1935), Mitchell (1959), Daniel (1981), London and Anderson (1970), and Meister and Loftis (1977).

¹⁰ In 1964, on the eve of the Delano Grape Strike, the value of California farm products sold was \$3.5 billion, 66% of which was sold by 7,000 farms. Just 7% of the farms farmed 78% of the 37,000,000 acres of agricultural land in the state or 33% of all farmland in the United States (U.S. Bureau of the Census 1968).

injunctions, jailings, and violence, while state and federal officials could be counted upon to find alternative, usually immigrant, sources of labor.

Despite these challenges, since the 1880s, farmworkers made repeated attempts to organize. When ethnic minorities took the initiative, they usually formed ethnic labor associations. Chinese (1880s), Japanese (1900s), Mexican (1920s), and Filipino (1930s) farmworkers all tried one version or another, some of which were more successful than others, but all of which ultimately failed to yield the desired protection. Union attempts, on the other hand, with the exception of CAIWU, were usually made when or where Anglos constituted a major proportion of the workforce—such as IWW and AFL attempts to organize “fruit tramps” around 1910, CIO and Teamster attempts to organize “Okies” in the 1930s, and the AFL’s National Farm Labor Union (NFLU) organizing attempt in the 1940s. None of these efforts succeeded in making lasting improvements in the lives of the workers, in building a stable membership base, or in getting union contracts.

In 1959, the year the AWOC was chartered, some 350,000 workers were employed in California agriculture, of whom some 90,000 were braceros—Mexican nationals imported as harvest hands under treaty between the United States and Mexico since 1942 (Sosnick 1978). The bracero program was at risk because of lobbying by farmworker advocacy groups (Proceedings, National Sharecropper’s Fund 1957; National Advisory Committee on Farm Labor 1967; Craig 1970; Meister and Loftis 1977; Mitchell 1979), diminished demand for braceros outside California due to mechanization of cotton and tomato harvesting (Craig 1970; Martin and North 1984), opposition by Midwestern farmers to “privileges” afforded California agriculture (Craig 1970; Martin and North 1984), and concerns about domestic unemployment following the 1958 recession (Carmines and Stimson 1989). In August, 1959, the Senate Subcommittee on Migratory Labor was formed, and on Thanksgiving day, 1960, Edward R. Murrow aired a graphic documentary on migrant farmworkers titled *Harvest of Shame*.

For these and other reasons, AFL-CIO President George Meany chartered the AWOC in February 1959 and shortly thereafter allocated it an organizing budget of \$250,000 a year (Taylor 1975; Meister and Loftis 1977; Mitchell 1979). Meany was influenced by the National Advisory Committee on Farm Labor, a citizen’s group chaired by Eleanor Roosevelt with ties to church leaders, labor leaders, liberals, and Democratic politicians (IUD Bulletin 1959; Taylor 1975; Meister and Loftis 1977; Mitchell 1979; Zieger 1987). Some labor leaders also believed that organizing agricultural workers could contribute to “curbing the political power of agribusiness” (Gorman 1959, p. 35) as “some of the most reactionary forces in the United States” (IUD Bulletin 1961, p. 15). Finally, because of their rivalry for leadership of the Federation, Meany hoped to avoid

giving the “organizing issue” to Walter Reuther, president of the United Autoworkers and vice president of the AFL-CIO (Meister and Loftis 1977; Mitchell 1979; Barnard 1983; Lichtenstein 1968).

Early in 1961, as the AWOC seemed to be making progress, the Teamsters also declared an interest in farmworker organizing, signing a nominal contract with a lettuce grower whom the AWOC had struck (Taylor 1975; Meister and Loftis 1977; Jenkins 1985; Daniel 1987).¹¹ Then, in 1962, during a hiatus in AWOC organizing, the independent National Farm Workers Association (NFWA), led by former community organizer Cesar Chavez, was founded and began organizing. By early 1966, however, it was the NFWA—not the AWOC or the Teamsters—that won the first genuine union contract in California agriculture. By 1970, as the United Farm Workers Organizing Committee (UFWOC, AFL-CIO), this union had brought the entire California table grape industry under contract, unionized some 70,000 workers, and achieved clearer success than any farm labor organizing effort in U.S. history.

THE AWOC AND THE UFW, 1959–66

The following account compares choices made by AWOC and UFW leadership at three critical junctures: their organizational foundings (1959–65), the Delano Grape Strike (1965–66), and the Schenley Boycott (1966). After comparing their strategy—emphasizing differences in targeting, timing, and tactics—I contrast the leadership and organization of the two groups, demonstrating the difference in their strategic capacity.

Organizational Foundings, 1959–65

AWOC’s strategy unfolded in two phases under two different directors. When chartered by the AFL-CIO in 1959, the AWOC was charged to organize farm workers to improve their wages, hours, and working conditions and support efforts to repeal the bracero program (Taylor 1975; Meister and Loftis 1977). AWOC’s strategy was to mobilize the workers, motivate them to pressure their employers, and get contracts. Although the struggle over the bracero program unfolded in a national, political, and long-term arena, AWOC’s primary organizing mission was local, work centered, and short term.

¹¹ The Teamsters had been expelled from the AFL-CIO for corruption in 1957 and were thus under no obligation to support the AWOC. They signed a food processing contract to include field workers to prevent AWOC’s claims of a labor dispute from interfering with the employer’s access to braceros—who did the field work but were excluded from the contract because they were braceros.

The AWOC targeted workers based less on characteristics of the workers themselves than on workplace settings they deemed favorable for organizing. Although it operated for five years, AWOC's timing was short term and focused on visible results, quickly achieved. Its principal tactics—organizing at early morning pickup sites—or “shape-ups”—for wage strikes, exercising insider political pressure, and recruiting through labor contractors—extended a familiar repertoire of conventional labor union tactics that had served in quite different historical settings.

In contrast, during a brief volunteer phase in 1961 when it operated without “professional” leadership and found itself dependent for resources on its constituency, AWOC's strategy changed, anticipating that of the UFW by more than a year. AWOC activists targeted workers most likely to provide a long-term organizational base, rather than those whose workplaces seemed to offer short-term advantages. They took a longer-term time perspective, and their tactics mobilized around the broad range of farmworker needs, not only their work situation. As is shown below, this strategic consistency, the variation observed within it, and its ineffectiveness can be explained by changes in AWOC's leadership, the structure of its organization, and how these influenced its strategic capacity.

During the first phase (1959–61), under the direction of Norman Smith, the AWOC targeted workers who gathered daily for early morning shape-ups because they were “easier” to organize (London and Anderson 1970; Jenkins 1985; Anderson 1996). These were mostly white single men, casual day laborers in very seasonal crops, a rapidly disappearing remnant of “dust bowlers” in an increasingly Mexican workforce (Metzler 1964; California Assembly Committee on Agriculture 1969; Villarejo 1997). AWOC's tactics were to conduct leafleting campaigns among these workers and to call strikes to raise wages. During its first 18 months, the AWOC led more than 150 strikes, some of which yielded temporary wage increases, but it failed to produce stable membership or a union contract (Meister and Loftis 1977; Jenkins 1985).

In early 1961, the AWOC shifted its tactics but continued targeting workers based on short-term political advantages and an “insider” political repertoire. AWOC leadership persuaded itself the new Democratic administration would enforce bans against use of braceros as strikebreakers (Taylor 1975). Targeting lettuce growers who were major users of braceros, the AWOC mobilized unemployed domestic workers to picket them, persuaded the few domestic workers employed there to walk out, claimed the existence of a strike, and called on Labor Secretary Goldberg to order the braceros out (London and Anderson 1970; Taylor 1975; Meister and Loftis 1977; Anderson 1996). Despite costly contests in local courts, the effort met with initial success. It backfired, however, when organizers

used violence to create a threat to the safety of the braceros so the Mexican consul would insist on their withdrawal. AWOC leaders were jailed and ended up with fines and legal bills approaching \$50,000, which the AFL-CIO had to pay. When AFL-CIO auditors checked the books, they also found inflated membership reports. This offense, combined with a general lack of results and the existence of jurisdictional issues within the AFL-CIO, was enough for Meany to fire AWOC Director Smith, close down the AWOC, and transfer its members to other unions (London and Anderson 1970; Taylor 1975; Meister and Loftis 1977).

Significantly, however, during the nine months the AWOC was officially shut down—and a new leadership team emerged, operating within a different organizational structure—a new strategy also emerged based on mobilizing volunteer support among farmworkers, farmworker advocacy groups, and within the labor movement itself (Anderson 1961; London and Anderson 1970; Meister and Loftis 1977; Anderson 1996). Among farmworkers, AWOC volunteers targeted stable resident Mexican families, organizing them to create “area councils” around a number of community concerns including housing, health care, and so on. They convened farmworkers and supporters for a widely attended organizing conference in December 1961 and organized a very dramatic appeal at the national AFL-CIO meeting early in 1962. The ironic outcome was that the AFL-CIO refunded the AWOC—but also hired another “professional” director who returned to the old strategy.

The new director, A. C. Green, targeted labor contractors whom the AWOC could picket. Contractors were middlemen paid by growers to recruit workers, who often took advantage of their role as “brokers” to cheat workers and growers. When the picketing worked, contractors signed agreements making their workers AWOC members and deducted dues from the workers’ pay. The intent was to produce immediate membership growth and, using contractors to control the labor supply, to get better terms from the growers. The membership grew, and the AWOC eventually signed 136 contractors (London and Anderson 1970; Taylor 1975; Meister and Loftis 1977; Jenkins 1985). But there were hundreds more it could not sign up, the contractors had no control over what growers chose to pay, and workers learned an AWOC contract meant paying dues for nothing in return (London and Anderson 1970; Taylor 1975; Anderson 1996). The only positive, if unintended, result of the focus on contractors was that Green hired two Filipino labor organizers, including Larry Itliong, with ties to Filipino labor contractors or crew leaders. Filipino crew leaders had traditionally negotiated with growers on behalf of crews of skilled workers, serving as advocates as well as supervisors. The initiative of these crew leaders actually started the Delano Grape Strike

(see below). By September 1965, on the eve of the grape strike, the AWOC had spent over \$1,000,000 and had failed to create a genuine membership base or to sign a single contract with a grower.

The FWA, by contrast, which began organizing in 1962, developed a strategy of organizing a community of workers, developing mutual benefits to strengthen it, and then putting pressure on the employers to get contracts (Nelson 1966; Levy 1975; Taylor 1975; Meister and Loftis 1977; Padilla 1999). The FWA leadership believed a critical error farmworker organizers had made in the past was in trying to strike and organize at the same time, the most recent evidence of which was the AWOC debacle in Imperial Valley the previous year (Taylor 1975; Levy 1975; Meister and Loftis 1977; Hartmire 1996; Drake 1997). Unlike the AWOC, the NFWA saw its strategy unfolding in a community, statewide, and long-term arena.

The FWA targeted Mexican resident farmworker families who were the growing part of the workforce (Metzler 1964; California Assembly Committee on Agriculture 1969; Jenkins 1985). A major center of this workforce was in Delano, the heart of the table grape industry, which provided one of the longest periods of employment in California agriculture. The FWA's tactics were to build a statewide association, blending community organizing techniques with a mutual benefit society and an ethnic labor association. Its timing extended over the five-year period the leaders believed would be required before they would be ready to confront the growers (Taylor 1975; Levy 1975; Drake 1997; Padilla 1999). The organizing began in the spring of 1962 with a statewide house meeting drive in farmworker communities, leading to a fall founding convention of the *Asociacion de Campesinos* or Farm Workers Association. "Association" was selected to avoid turning away workers with negative experiences in earlier unionization attempts or provoking a premature reaction from the growers. "Campesino" was descriptive of the Mexican peasantry, whose movement since the Revolution was evocative of land, dignity, and resistance (Levy 1975; Taylor 1975; Meister and Loftis 1977; Medina 1998; Padilla 1999).

Within three years, the FWA had established a small death benefit, a social service program, credit union, newspaper (*El Malcriado*), its own flag, small treasury (\$1,700), two paid staff, 1,500 members, and a capacity to engage in rent strikes, small work stoppages, and the like (Taylor 1975). Although largely self-sufficient to this point, in the summer of 1965, before the strike began, the NFWA applied for an Office of Economic Opportunity (OEO) grant of \$500,000 to develop a range of cooperative community services for farmworkers (Taylor 1975). As we will see below, by the time the grant was awarded, circumstances had changed so dramatically that the NFWA decided to turn it down.

I now show how variation in the composition of the leadership and the organizational structure of the AWOC and the UFW not only yielded the difference in strategy described above, but a sharp difference in the underlying capacity to develop effective strategy—a point which becomes increasingly clear as events unfold.

Leadership

Salient differences among the leaders, staff, and volunteers of the two efforts become quite clear upon review of table A1, in the appendix, which summarizes biographical data including name; age at the time participant undertook farmworker organizing; race or ethnicity; religion; regional background; position; family background; family status; education, work, and organizing experience; work commitment; network affiliation; and repertoires of collective action. Family status is distinguished as married (M), single (S), children (C), divorced (D), and whether the spouse was active in the organizing effort (A). Interest in organizing of farmworkers is distinguished as to whether it was professional (an assignment), vocational (a mission), or personal (about one's own life). The appendix offers brief life narratives for key actors.

AFL-CIO president George Meany, organizing director John Livingston, and AWOC directors Norman Smith and A. C. Green developed the AWOC strategy. All were white men, age 52 or over, with extensive union backgrounds and experience at “insider” politics. As table A1 shows, the men whom Meany and Livingston chose to lead the AWOC were of a generation of “union men” who valued “legitimate” ways to do union work and had little understanding of workers different from themselves or of a public whose support they would need. They had no biographical experience of the farmworker community, few sociocultural networks reaching beyond their milieu (much less into the farmworker community), and tactical repertoires learned by organizing people like themselves in circumstances far different from those they now faced. Smith, who had done no union organizing for 18 years, was chosen because of a relationship with Livingston going back to UAW organizing in Flint, Michigan, and because more likely candidates such as Ernesto Galarza (see below) or Clive Knowles were associated with rival international unions, neither of which Meany wanted to offend (Anderson 1996).

Among those who worked with the AWOC but remained outside the leadership circle, however, were the 17 organizers whom the AWOC hired, the farmworkers it was trying to organize, and community supporters. The life experience, networks, and tactical repertoires of many of these people—such as Father Tom McCullough, Henry Anderson, Dolores Huerta, Andy Arellano, Cipriano Delvo, Raul Aguilar, and Larry Itli-

ong—linked them not only to the farmworker community, but to religious, student, and liberal groups as well. Most were younger, many were Mexican American or Filipino, and a few were women. Dr. Ernesto Galarza, the Mexican academic who had worked with the NFLU since 1949, served as Smith's assistant for 6 months, until October 1959 when he resigned over jurisdictional concerns. His main role had been to instruct Smith and the organizers in how to document abuses of the bracero program, a mission he had pursued for the previous eight years (Anderson 1996). Since this staff was not party to AWOC decision making, however, no one making strategic choices had ties to the farmworker constituency or "local knowledge" of the conditions about which they were making choices. As a result, the organizing repertoires decision makers brought to this new situation constrained more than they enabled. The personal commitment motivating AWOC's leadership to find ways to make the effort successful was also very limited. Meany's commitment to the AWOC was minimal. His decisions about how much support to give were the result of political pressure from farmworker advocacy groups, their liberal allies, and his rival, Walter Reuther. For Livingston, Smith, and Green, organizing farmworkers was an assignment, not a mission.

The composition of the FWA leadership team was far different from that of the AWOC, combining insiders and outsiders, those with strong and weak ties, and a diversity of salient repertoires. The FWA strategy was developed by leaders who were Mexican and Mexican American men and women mostly under 35, whose lives were rooted in the farmworker community but extended well beyond it—such as Cesar Chavez, Dolores Huerta, and Gilbert Padilla. They worked in collaboration with two white clergymen affiliated with the California Migrant Ministry (CMM), Chris Hartmire and Jim Drake, both of whom were under 30 with middle-class backgrounds but who, as a result of their seminary experience, had come to share a vocation to improve the lives of farmworkers. FWA leaders thus drew on life experience that combined "local knowledge" of the farmworker world with experience in military service, college, small business, and professional organizing. The clergy brought "local knowledge" of religious groups and insight into the middle-class support constituency from which they themselves were drawn. Many of the leaders benefited from the insights of "borderland" experience as young Mexican Americans growing up in the 1940s. The leadership of the FWA also had a deep personal interest in finding ways to succeed at what was a personal mission. Not only had they come from farmworker backgrounds, they had given up secure jobs and other opportunities to risk building a new organization from the ground up. This combination of personal, vocational, and professional interests infused the effort with powerful motivation to develop a strategy that would make it work.

The range of sociocultural networks with which they were affiliated extended from the farmworker community into the worlds of community organizing, Mexican American activists, religious groups, and liberal circles throughout California. These networks combined strong ties to the farmworker community and the religious community with weak ties to many groups who would play important roles in the organizing (Granovetter 1973).

Finally, their tactical repertoire grew out of their experience and training as professional community organizers, particularly within the Spanish-speaking community. But they also drew on some union experience (Huerta and Chavez), Catholic retreat training (Chavez), and electoral experience (Chavez, Huerta, Padilla). The house meeting drive and founding convention used to kick off the FWA, for example, was a tactic adapted from numerous local Community Service Organization (CSO) organizing drives to statewide purposes.

Neither the AWOC nor the FWA “had to” pursue the strategies they did—a fact demonstrated in AWOC’s case by the alternate approach taken by volunteers during the organizational hiatus. To learn why the experience of its leaders imposed “limits” on the effectiveness of its strategy—and why the experience of FWA leaders was such a fruitful source of strategy—we turn to the organizational setting within which strategy was developed.

Organization

While the FWA’s deliberative structure was anchored in regular board meetings and inclusive strategy sessions, AWOC’s deliberative structure provided no focal point for creative discussion. Within the AWOC, not only were there no advisory councils or farmworker committees, but staff meetings were irregular, lacked agendas, and were venues for announcements, not strategic reflection—especially after Galarza left. Either Director Smith or Director Green was in charge.¹² They decided what to do. The organizers’ job was to do what they were told (Anderson 1961; London and Anderson 1970; Meister and Loftis 1977; Anderson 1996). This severely limited opportunity for diverse perspectives to be heard, for reflection, and for learning. In the FWA, on the other hand, regular board meetings and strategy sessions anchored a deliberative process, which included a far wider leadership group (including Drake and Hartmire) in

¹² Although members of the National Advisory Committee on Farm Labor shared responsibility for the formation of the AWOC—and were later to support the UFW—the AFL-CIO insulated them from any input into AWOC’s strategy.

frequent discussion of the choices facing the organization (Nelson 1966; Hartmire 1996; Drake 1997).

Resources flowed from the top down within the AWOC, motivating the development of strategies that would satisfy those at the top, while the FWA's financial and human resources flowed upward, motivating the development of strategies that could yield the needed resources. For the AWOC, the main organizing resource was money to pay staff, cover organizing expenses, and maintain offices. The fact that dues were not a significant source of income meant there was limited motivation to create a financial base among workers. The AWOC also had little motivation to generate volunteer participation from farmworkers or supporters, relying on paid staff to do its work (Anderson 1961; London and Anderson 1970; Meister and Loftis 1977; Anderson 1996). The FWA's financial resources, on the other hand, were based on membership dues of \$3.50/month per family, which were collected in cash by local representatives. Dues entitled member families to a small death benefit, social services, membership in the credit union, and a newspaper. Chavez's personal savings and individual contributions from supporters supplemented these resources. The UFW's human resources included one intermittently paid staff member (Chavez earned \$50 a week), which grew to three staff members by 1965 (with the addition of Huerta and Bill Esher, newspaper editor). In addition, the CMM paid Padilla part time and offered the services of Drake and Hartmire as needed and available (Taylor 1975; Smith 1987; Hartmire 1996; Drake 1997; Padilla 1999). Most of the organization's work, however, was done on a volunteer basis by board members, representatives, and other activists—resting on maintenance of the extraordinary level of motivation that underlay the whole undertaking.

The accountability structure of the AWOC was based on a chain of command that precluded input from those below the top levels at which strategy was discussed, while that of the FWA was based on responsiveness to the farmworker community and to those doing the daily work of the organization. AWOC's bureaucratic command structure was viewed as the only source of strategic legitimacy within the organization (Anderson 1996). As an organizing committee of the AFL-CIO, Meany appointed its directors, on the recommendation of Livingston, and the directors, in turn, hired the staff who reported to them. There was no advisory board of supporters, accountability mechanism to farmworkers, or role for farmworker leadership within this chain of command (Anderson 1961; Anderson 1996). Not even Livingston was permitted to sit in on Executive Council settings in which Meany formulated strategy (Reuther 1976). The FWA, on the other hand, was built around a six-person executive board (Chavez, Huerta, Padilla, Orendain, Terronez, and Hernandez) elected by farmworker delegates at the founding convention. It was led by a full-time

president, part-time volunteer officers, and a network of appointed local representatives in each community (Taylor 1975; Padilla 1999). Chavez, Huerta, and the others had also “selected themselves” for this mission.

Strategic Capacity

Differences in deliberative processes, resource flows, and accountability structures interacted with differences in the composition of leadership teams to create far greater strategic capacity for the UFW than for the AWOC. AWOC’s extremely narrow deliberative process limited the quality of leaders’ thinking by excluding the diverse—and more salient—perspectives of others in and around the AWOC. The top-down resource flow meant not even AWOC directors had autonomy to develop strategy that assigned first priority to the mission of organizing farmworkers. This is illustrated when Meany shut down the AWOC in 1961 and assigned Green to do political work for the better part of 1962, his first year as AWOC director. The fact that AWOC’s leadership was not accountable to farmworkers meant the concerns of Meany and Livingston would always carry the day. “Getting quick results” or “building up the membership numbers” had a far greater influence on Smith and Green than would have been the case if they had to deal with an organized farmworker constituency whose reactions they had to consider as well. Ironically, the top-down accountability structure and resource flows yielded little motivation to develop organizing strategies that could produce resources—financial resources through worker dues or contributions of supporters, or human resources such as volunteering by farmworkers and others. AWOC leaders had the resources to keep making the same mistakes, as long as the people at the top were satisfied.

FWA leadership, on the other hand, made the most of its capacity because of organizational arrangements it made. FWA’s deliberative process assured extensive and diverse input. Inclusion of the clergy infused the conversation with input from those not in a direct chain of command and thus more likely to argue their own point of view. FWA leaders drew heavily—but not exclusively—on the farmworker community for both human and financial resources. Need for these resources motivated development of strategy that would successfully expand its membership and base. FWA’s accountability structure also tied its leadership to those whose active support was needed if the enterprise was to succeed. Although it was based on election by a farmworker convention, on a day-to-day basis, as members of a “committed band,” leaders were accountable to each other, including Chavez and the officers, through board meetings. The leaders were thus motivated to devise effective strategy and enjoyed the autonomy to act on strategy they devised—single-mindedly pursuing

a farmworker organizing agenda. This had the consequence that the FWA would make its allies uncomfortable from time to time but never find itself so dependent upon any one of them that a supporter's priorities would supplant its own (Levy 1975; Daniel 1987; Cohen 1995; Hartmire 1996; Chatfield 1996).

The contrast between AWOC and FWA leadership in terms of access to salient information, heuristic facility, and motivation could not have been greater. The diverse experience of FWA leadership, working in a productive organizational setting, allowed it to develop strategy more effectively than the AWOC. Although the AWOC had access to far richer leadership resources than it utilized, it structured itself to preclude their engagement in the development of strategy—which was developed by those least well equipped to do so. The differences in outcomes achieved by the two organizations during this founding period were more than the result of “mistakes” on the part of the AWOC and “brilliance” on the part of the UFW. As FWA leaders drew on their strategic capacity to devise the targeting, timing, and tactics with which it could achieve its goals, what emerged was something new: a combination ethnic labor association, mutual benefit society, and community organization. And, as we will see below, when these tactics did not work, the leaders had also developed the capacity to change.

The Delano Grape Strike, 1965–66

The following analysis of the Delano Grape Strike shows the value of accounting for outcomes not only in terms of a specific strategy, but in terms of an underlying, and developing, strategic capacity. Although the bracero program had finally been phased out in 1964, altering the organizing environment, a grape strike was not on the strategic agenda of either organization.¹³ When it occurred, AWOC leadership viewed it as a short-term minor diversion from its major focus at the time, a joint campaign with the Teamsters to organize citrus workers in another part of the Valley (Taylor 1975; Meister and Loftis 1977). NFWA leadership on the other hand viewed it as a risky long-term diversion from their plan that had called for two more years of organizing without a major strike (Nelson 1966; Levy 1975; Taylor 1975; Hartmire 1996; Drake 1997; Padilla 1999). The different ways leaders of these two organizations responded to this unanticipated, but “eventful” (Sewell 1996), “focal moment” (Lofland

¹³ Many of the single men who had worked as braceros returned to the United States—but as legal immigrants were not barred from organizing. Their experience had convinced many to become strong UFW supporters in vegetable and citrus industries where they came to be employed.

1996) set the NFWA on the path to success and the AWOC on the path to dissolution.

Strategizing the Strike

The 1965 grape strike was the consequence of an initiative by AWOC organizer Itliong and Filipino crew leaders trying to raise wages of their Delano grape crews from \$1.20/hour to \$1.40/hour, the minimum wage that growers had been required to pay braceros. These crews comprised some 800 of the 3,500 grape workers in Delano, but they included the most skilled workers. Forced to respond, but not wanting to alienate the Filipino crew leaders unnecessarily, AWOC leadership agreed to support a strike for a wage increase to be quickly won or lost depending on whether Filipino grape workers who walked out stayed out. AWOC's tactics focused on providing meals for strikers, housing when needed, and a modest strike benefit. AWOC staff was personally unaffected, remaining on salary and expenses as before. The AWOC maintained stationary picket lines at the 10 companies that it had struck, mainly to keep its members occupied and to make sure none of them returned to work (Nelson 1966; Taylor 1975; Meister and Loftis 1977).

Meanwhile, the leadership of the NFWA struggled to respond strategically to a strike called in its midst without its knowledge or consent. The Filipino strike had struck a nerve with the 2,500 Mexican workers, who had gone many years without a substantial wage increase and among whom the NFWA had several hundred members (Nelson 1966; Taylor 1975). NFWA leadership had just concluded a minor but successful strike action of its own but was dubious of its capacity to manage a major sustained strike (Nelson 1966). On the other hand, sensing an opportunity and willing to take some risk, NFWA leaders decided to test support among Mexican workers by mobilizing them for a public strike-vote meeting (Taylor 1975; Nelson 1966; Drake 1997; Hartmire 1996; Padilla 1999). Drawing on Mexican tradition, the NFWA encouraged attendance—and a sense of courage and commitment—by meeting in the hall of Our Lady of Guadalupe Church in Delano, the religious center of the community, on September 16, Mexican Independence Day. NFWA leaders had decided to risk leading a strike if the turnout were substantial and the workers would agree to three strategic conditions—conditions that indicated how the NFWA had already begun to redefine its strategy (Nelson 1966; Levy 1975; Taylor 1975; Meister and Loftis 1977; Hartmire 1996; Drake 1997).

First, unlike the AWOC, the NFWA had no strike fund and could not offer strike support, so everyone would have to share the risk. Although this was not a viable option for the AWOC because of the isolation of the leadership from its constituency, it was for the NFWA because of the

social capital it had built among the workers, the level of commitment among the leadership, and the confidence it had developed in learning how to survive by relying on volunteer resources (Taylor 1975).

The second condition was that the strike would be nonviolent. This was new to the farmworker community and to agricultural strikes in general. Although Chavez had long been interested in Gandhi, reframing the strike as a "nonviolent struggle" would also help the NFWA find support to sustain it from church groups and others with whom it already had relationships (Hartmire 1996). The value of identification with the Civil Rights movement had become clear in a rent strike the UFW led earlier in the summer and that came to the attention of California civil rights groups who supported it (Chatfield 1996; Padilla 1999). Just five months earlier, the whole country had observed the impact of Dr. King's Selma to Montgomery March and just five weeks earlier that of the Watts Riot.

Third, the strike would be for union recognition, not only for the wage increase the AWOC was seeking (Taylor 1975). This framed the strike in terms that would remind unions and others, particularly the churches, that farmworkers remained outside the National Labor Relations Act (NLRA), had to strike for recognition, and therefore were deserving of their support.

The 1,200 enthusiastic Mexican workers who attended the meeting voted overwhelmingly to accept these conditions and to go on strike (Taylor 1975; Meister and Loftis 1977; Levy 1975; Medina 1998; Padilla 1999). This was the beginning of the grape strike that unfolded less as a minor labor dispute (which it was for the AWOC) and more as the first step in the birth of a farmworkers movement (which it was for the NFWA).

The Strike Unfolds: Strategic Choices

As the strike unfolded over the next few weeks, NFWA leadership drew on its strategic capacity to adapt successfully to new challenges with a new urgency of timing and with new tactics. These choices, in turn, began to reshape the contours of the organization itself and its leadership in ways that deepened the strategic capacity on which it could draw to face subsequent challenges. The extent of the strategic reframing, which had already occurred, became clear as early as October 1965 when the NFWA got the news it had been awarded \$270,000 of the OEO grant it had applied for and turned it down. The NFWA would not abandon its role in the strike, as receiving the government funds would have required (Levy 1975; Taylor 1975; Chatfield 1996). AWOC leadership, on the other hand, drawing on very limited strategic capacity—but more abundant material resources—grew increasingly isolated from its own strike leaders and al-

lies as time passed (Nelson 1966; Taylor 1975). The different choices NFWA and AWOC leaders made in dealing with the same situation show the influence differences in the leadership and organization of the two groups had on their capacity to devise effective strategy.

While the AWOC limited the effectiveness of its strikers by assigning them to picket only farms on which they had been employed, the NFWA learned how to lead a strike of 2,500 workers with just 200 pickets. Since the AWOC hoped to win the strike by keeping its own members from returning to work, it established "stationary picket lines" outside the vineyards in which they had worked, mainly to keep them occupied and vigilant (Nelson 1966). On the first morning the NFWA joined the strike, however, only 100 to 200 activists reported to begin picketing. Realizing an effective strike would require the participation of many more workers than these 200, the NFWA devised the "roving picket lines" tactic (Nelson 1966; Taylor 1975; Drake 1997). Car caravans of pickets arrived at grape fields waving flags and banners, called the workers out of the fields, and then moved on to the next location. Although 2,500 of the 3,500 Delano grape workers eventually joined the strike and refused to go to work, most left the area to find work elsewhere rather than joining the picket lines (Nelson 1966; Taylor 1975; Meister and Loftis 1977). With the "roving picket line," a relatively small core of NFWA activists could sustain the strike longer—and with less money—than anyone expected. They became the core of a full-time activist cadre, many of whom would become organizers (Brown 1972). The AWOC rejected NFWA proposals for joint strike committees or picket lines, leaving the NFWA free to develop its own tactics (Nelson 1966; Taylor 1975; Drake 1997; Padilla 1999).

Unlike the AWOC, the NFWA had to generate its own strike fund, so it learned to use its networks to begin mobilizing public support. Hartmire recruited clergy delegations to come to Delano, to see the conditions "first hand," and to return home to raise food and money (Nelson 1966; Taylor 1975; Hartmire 1996; Drake 1997). Civil rights groups such as the Bay Area Friends of Student Nonviolent Coordinating Committee (SNCC) mobilized food caravans to Delano (Taylor 1975). A fund-raising speech by Chavez at the University of California, Berkeley, not only netted \$6,000, but encouraged students to volunteer in Delano as well (Nelson 1966; Taylor 1975). In contrast, the AWOC rejected all offers of "outside" support, except those coming through "legitimate" labor channels. The AWOC also rejected the NFWA's proposal to create a joint strike fund (Nelson 1966; Taylor 1975; Padilla 1999). Although AWOC leadership did not think it needed "outside" support, this decision isolated it from a public that had become increasingly interested in the strike (Taylor 1975).

The NFWA learned to turn local injunctions intended to cripple effective picketing into opportunities for civil disobedience, which would gen-

erate wider public support for the strike. The AWOC relied on the advice of its AFL-CIO lawyers to avoid costly legal entanglements (Taylor 1975). The NFWA, in contrast, based its response on the advice of volunteer civil rights lawyers, a far wider tactical repertoire, and greater willingness to take risks. As grower recruitment of strike breakers from outside the area grew more effective, workers no sooner left the fields than others would take their places, requiring the NFWA to encourage them to leave as well. When the local sheriff barred strikers from shouting "huelga" (Spanish for strike) to workers in the fields, the NFWA planned a well-publicized arrest of 44 persons for exercising "free speech" (Nelson 1966; Taylor 1975; Meister and Loftis 1977; Drake 1997). The fact that 11 clergymen were among those arrested evoked similar scenes of civil disobedience in the South. Similar tactics kept the strike in public view, helping to generate the resources to sustain it (Taylor 1975). One result of growing public attention was a December 1965 visit to Delano by Walter Reuther who pledged \$5,000 per month in support to the strike. Significantly, after intensive lobbying by NFWA supporters, Reuther split the money between the AWOC and the NFWA, giving the independent NFWA the first substantial labor support it had received (Nelson 1966; Taylor 1975; Schrade 1984; Chatfield 1996; Drake 1997). This was a rebuke to AWOC leadership and a sign some labor leaders were taking another look at the Delano strike, particularly those critical of Meany's leadership of the Federation. It also showed the NFWA had learned to see "opportunity" in the conflict between Reuther and Meany that it could turn to its advantage.

Ethnic identity had been central to the NFWA organizing strategy since its founding convention, in sharp contrast to the AWOC (Levy 1975; Taylor 1975). While there were solidaristic benefits to this approach among its membership, it was also a way the NFWA could reach out to farmworkers who knew less about the benefits it offered but understood that it was an effort of the "Mexican people" to help themselves (Medina 1998). The rich Mexican cultural tradition also provided "moral resources" strikers drew upon to sustain their motivation. Roman Catholic masses celebrated by "huelga priests" affirmed values of sacrifice and solidarity. Traditions of mutuality among extended families modeled the mutuality at the core of the striker community, and Mexican history came alive as slogans appeared on walls that read: Viva Juarez! Viva Zapata! Viva Chavez!

But it also had benefits in the country as a whole. The systematic discrimination to which Mexicans had been subjected in the Southwest was a story not well known by the rest of the country, but the NFWA leadership's recognition of the public support developed by the Civil Rights movement suggested this might be a story that the rest of the country

could be told. It would help explain the dire circumstances in which farmworkers had come to live, while distinguishing the farmworker struggle from “just another strike” and the NFWA from “just another union” (Levy 1975; Taylor 1975; Hartmire 1996). It would also draw support of urban Mexican Americans. Although public support for civil rights offered farmworker organizers a new opportunity, it was not an opportunity AWOC leadership recognized. Trapped within a reality defined by its own leadership and organizational structure, it insisted on its identity as “just another union” and the grape strike as “just another strike” (Nelson 1966; Taylor 1975).

Strategic Capacity

The development of a “strike community” was the most far-reaching consequence of the choices NFWA leaders made about how to conduct the strike under severe financial constraints. By transforming the organization and its leadership, this increased its access to salient information, enriched its heuristic facility, and deepened its motivation—and created a critical core for the emergence of a social movement (Kim and Bearman 1997). The AWOC, on the other hand, relying on AFL-CIO financial support adequate for its limited objectives, saw no need to change. To retain the support of the strikers, NFWA leaders believed they had to share the strikers’ level of sacrifice, which meant strike benefits of \$1.00 per week (later \$5.00 per week) and food orders from a strike “store” (Nelson 1966; Taylor 1975). Because of their depth of personal commitment, NFWA officers moved to Delano, became full-time volunteers, and supported themselves as strikers—which, in turn, deepened their commitment to winning the strike and made it easier for them to claim similar levels of commitment from others (Brown 1972). There were three important consequences to this choice.

First, because the cost per person was so low (food, a bed, \$1.00 per week), the NFWA could relatively easily add full-time volunteers, and it began accepting students and religious activists who came to Delano to join the strike on the same terms (Nelson 1966). By enabling large numbers of people to volunteer, the NFWA developed a new talent pool on which it could draw for the myriad new responsibilities that had begun to emerge. Expansion of this cadre—for which the roving picket line had served as a core—made it possible for the NFWA to field large numbers of full-time “troops” for strike, boycott, and political activities.

Second, in a setting as “open” as the NFWA, bringing in new people facilitated the emergence of new leadership that, in turn, expanded, enriched, and altered the composition of the original NFWA leadership group, in ways that further enhanced its strategic capacity. As shown in

table A1 below, the volunteers included young farmworker leaders who emerged from striking families such as Eliseo Medina, Marcos Munoz, and Maria Saludado. They would serve as picket captains, organizers, boycott organizers, and some later rose to leadership positions in the union. Of some 40 boycott cities operating at the peak of the grape boycott, some 35 were led by new farmworker leaders (Brown 1972). Another source of people was students who had been involved in the Civil Rights movement who came to Delano as volunteers. Individuals such as Luis Valdez, Jessica Govea, and I served in a wide variety of roles as organizers, boycotters, administrators, and so on (Nelson 1966; Taylor 1975; Levy 1975; Mathiessen 1969). A third font of volunteers were religious activists, such as Leroy Chatfield, inspired by the Civil Rights movement, recruited by the CMM, or motivated by Vatican II (Hartmire 1996; Chatfield 1996; Drake 1997). Finally, talented young lawyers, such as Jerry Cohen, were attracted—drawn to public service but wishing to practice more “political” law than was possible through legal services corporations (Levy 1975; Cohen 1995; Chatfield 1996).

Third, these choices led to the emergence of a “charismatic community” based on “vows of voluntary poverty” that shared an almost religious commitment to winning the strike (Nelson 1966; Taylor 1975; Meister and Loftis 1977; Daniel 1987). Amplified by solidaristic tactics needed to sustain this level of commitment, such as Chavez’s 28-day fast in 1968, this community became a “crucible” of cultural change (Turner 1966; Gamson 1991; Taylor and Whittier 1992; Peterson 1999), which transformed farmworkers into “chavistas,” supporters into “voluntarios,” the grape strike into “La Causa,” and Chavez into a legendary farmworker leader.¹⁴ This cultural dynamic was to infuse the UFW with significance for farmworkers, Mexican Americans, students, religious activists, and liberal Americans far beyond its political reach as a community organization or ethnic labor association, beginning to give it the impact of a genuine social movement.¹⁵

Summary and Comparison

The foregoing shows how the NFWA and the AWOC made strategic choices about how to deal with a grape strike neither group had “planned.”

¹⁴ Under these circumstances, talented leaders may be transformed into a symbol of this “new” community of identity—the source of “charisma” (Weber 1978; Durkheim 1915; Meindl 1989). Charismatic effects attributed to the leader, such as attracting followers, enhancing their sense of self-esteem, and inspiring them with a willingness to exert extra effort, may be the result of a kind of social contagion (House, Spangler, and Woycke 1991; Hollander and Offerman 1990; Meindl 1989).

¹⁵ The argument here is that social movements have a “symbolic reach” that extends their influence beyond those with whom it has direct personal interaction. “Cultural

The choices the NFWA made met its immediate needs but did so strategically—creating new opportunities for the organization to move toward its goals and at the same time expanding the access to salient information, heuristic opportunity, and motivation of its leadership. By December 1965, when the NFWA launched its first boycott, it had transformed its strategy. It saw itself not only as a community organization and ethnic labor association, but also as a farmworker civil rights movement. The choices the AWOC made also met its immediate needs but moved it in the opposite direction, toward isolation and dissolution.

The Schenley Boycott, 1966

As the harvest drew to a conclusion in November without wage increases or contracts with the growers, NFWA leaders had to face the new challenge of how to sustain a strike when there was no work—at least until pruning in January. Although NFWA strikers had followed grapes to produce terminals in Los Angeles and San Francisco since October, the main result had been injunctions against secondary picketing (Taylor 1975; Meister and Loftis 1977). Volunteer SNCC researchers, however, who had learned to investigate the financial connections of segregated institutions so they could be exposed and picketed in the North, began to investigate the growers. In December 1965, they discovered that Schenley Industries, which owned 5,000 acres of wine grapes in Delano, was a major liquor producer and distributor who marketed such well-known brands as Cutty Sark scotch whiskey (Taylor 1975; Drake 1997).¹⁶ Targeting Schenley for a boycott became a new option. Deciding the NFWA needed to “try something,” Chavez named Jim Drake of the CMM and Mike Miller, a San Francisco SNCC organizer, as co-coordinators of the boycott (*Boycott Newsletter* 1965; Drake 1997).

Timed to take advantage of the Christmas season and counting on support from religious, SNCC, Congress of Racial Equality (CORE), and Students for a Democratic Society (SDS) networks, the UFW dispatched a corps of student volunteers and strikers to major cities across the United

reception” theorists (Griswold 1987) would argue the symbols or cultural objects a social movement generates (including a symbolic leader) and disseminates can be appropriated by people who begin to make use of them in their own ways.

¹⁶ The significance of an environment conducive to strategic innovation—and the unpredictable way in which it often unfolds—is illustrated by the fact that the original impetus for research into Schenley was the rumor it was a Kennedy family investment. Although Joseph Kennedy had imported Cutty Sark during World War II, the rumor turned out to be false. It led, nevertheless, to consideration of the possibility of a boycott. March and Olsen (1976) note the importance of environments that permit whimsical paths to strategic innovation as employing the “technology of foolishness.”

States to organize picket lines of liquor stores in communities likely to respond (Taylor 1975; Levy 1975; Meister and Loftis 1977; Drake 1997). Their tactics were drawn more from a civil rights repertoire than a labor repertoire—they conducted a secondary boycott, asking consumers not to shop at stores being picketed until the stores removed Schenley products from the shelves (Taylor 1975; Drake 1997).¹⁷ Boycott organizers were also expected to raise operating costs in the cities in which they arrived, as the UFW could afford to send them no funds (Drake 1997). The AWOC was invited to take part in the boycott but declined because of legal concerns about NLRA prohibitions and because AFL-CIO Distillery Workers who represented Schenley winery workers vetoed labor support for the boycott (Taylor 1975; Levitt 1996).

Although there was an encouraging response from the public, the boycott wore on into January 1966 without visible result. While considering tactics to strengthen it, the leadership became concerned with how to keep workers from returning to Delano in the spring when work in the grapes resumed (Levy 1975; Taylor 1975). In a deliberative process to which the NFWA turned frequently when faced with critical choices, Chavez gathered a leadership group at a supporter's home in Santa Barbara to spend three days figuring out what to do. Besides Chavez, the strategy team included Huerta, Drake, Valdez, farmworkers Robert Bustos and Tony Mendez, myself, and others (Ganz 1994; Drake 1997). Perhaps the best way to give a sense of the creative process—and the interaction of people and ideas central to it—is to quote from my notes:

As proposals flew around the room, someone suggested we follow the example of the New Mexico miners who had traveled to New York to set up a mining camp in front of the company headquarters on Wall Street. Farmworkers could travel to Schenley headquarters in New York, set up a labor camp out front, and maintain a vigil until Schenley signed. Someone else then suggested they go by bus so rallies could be held all across the country, local boycott committees organized, and publicity generated, building momentum for the arrival in New York. Then why not march instead of going by bus, someone else asked, as Dr. King had the previous year. But it's too far from Delano to New York, someone countered. On the other hand, the

¹⁷ Boycotts had long been part of the standard union repertoire. In 1947, however, the Taft-Hartley Act made "secondary boycotts" illegal for organizations of workers covered by the NLRA. When unions ask consumers to shun buying anything in a store that sells struck products, it is a secondary boycott. The 1959 Landrum-Griffin Act extended the ban to the remaining "hot cargo" clauses, which had allowed union members to refuse to handle struck products; e.g., truck drivers, stevedores, supermarket employees. As a result, most union boycotts had become "pro forma" and usually only consisted of placing the boycotted product on an "unfair list" published in union newspapers (Miller 1961). Since 1956, however, the Montgomery Bus Boycott had revived it as an important tool of the Civil Rights movement.

Schenley headquarters in San Francisco might not be too far—about 280 miles which an army veteran present calculated could be done at the rate of 15 miles a day or in about 20 days. . . . But what if Schenley doesn't respond, Chavez asked. Why not march to Sacramento instead and put the heat on Governor Brown to intervene and get negotiations started. He's up for re-election, wants the votes of our supporters, so perhaps we can have more impact if we use him as "leverage." Yes, someone else said, and on the way to Sacramento, the march could pass through most of the farm-worker towns. Taking a page from Mao's "long march" we could organize local committees and get pledges not to break the strike signed. Yes, and we could also get them to feed us and house us. And just as Zapata wrote his "Plan de Ayala," Luis Valdez suggested, we can write a "Plan de Delano," read it in each town, ask local farmworkers to sign it and to carry it to the next town. Then, Chavez asked, why should it be a "march" at all? It will be Lent soon, a time for reflection, for penance, for asking forgiveness. Perhaps ours should be a pilgrimage, a "peregrinacion," which could arrive at Sacramento on Easter Sunday. (Ganz 1994)

The weaving together of diverse networks of people and ideas that characterized the planning of the march characterized preparations for its kickoff, as well. It was timed for March 17, the day after the Senate Subcommittee on Migratory Labor, with participation from Senator Robert Kennedy, was to hold hearings in Delano (Taylor 1975; Ganz 1994). This was an event organized with Reuther's help that would bring national media to Delano (Taylor 1975; Levy 1975; Daniel 1987). The march was targeted to accomplish three objectives: to win support for the strike by persuading workers along the march route to stay out of Delano when work began in the spring; to pressure Democratic Governor Edmund G. Brown, who was up for election that year and concerned about Mexican American voters, to intervene in the dispute; and to gain public support for the Schenley Boycott by demonstrating the injustice of the farmworkers' plight (Taylor 1975; Ganz 1994; Drake 1997). The march was led by a farmworker carrying a banner of Our Lady of Guadalupe, the patroness of Mexico, portraits of campesino leader Emiliano Zapata, and banners proclaiming "peregrinacion, penitencia, revolucion": pilgrimage, penance, revolution (Taylor 1975; Levy 1975). The marchers carried placards calling on supporters to boycott Schenley (Taylor 1975; Meister and Loftis 1977). The AWOC was again invited to participate, but Green declined, declaring the AWOC was involved in "a trade union dispute, not a civil rights movement or a religious crusade" (Taylor 1975, p. 153).

The march attracted wide public attention, particularly after television images of Delano police trying to block its departure evoked images of similar police lines in Selma, Alabama, the year before (Taylor 1975; Levy 1975; Meister and Loftis 1977). On April 3, 1966, one week before the march was to arrive in Sacramento, following mediation by a Bartenders'

Union representative, Schenley recognized the UFW covering the 500 grape workers it employed in Delano (Levy 1975; Taylor 1975; Levitt 1996; Hartmire 1996). The 82 original farmworker marchers reached Sacramento accompanied by 10,000 farmworkers and supporters.

This NFWA success dramatically altered the terms of the farm labor conflict and the arena within which the struggle was being waged. Because of the strategic capacity it had developed, the union had learned to keep finding ways to turn meager resources into effective economic weapons. The world of agribusiness turned out to be less monolithic than had been thought as divisions emerged between local agricultural corporations, whose entire business was farming, and national corporations, with major investments in brand names (Brown 1972; Taylor 1975; Meister and Loftis 1977). It made little sense for large corporations to risk compromising their brands for the sake of minor farming operations, especially when they had union contracts elsewhere (Levitt 1996). Farmworkers for the first time began to believe unionization was achievable, especially when they experienced the reality of the Schenley Contract: it not only increased wages, but eliminated the hated labor contractor system, provided for seniority and job security, and included a medical plan (Brown 1972; Taylor 1975; Medina 1998). The newly appointed AFL-CIO organizing director, William Kircher, fired Green, closed up most of AWOC's operations, and began negotiating a merger with the NFWA (Taylor 1975). The eventual agreement recognized the autonomy of a new United Farm Workers Organizing Committee (UFWOC) and committed AFL-CIO financial support of \$150,000 per year (Taylor 1975). The Teamsters discovered it had a common interest with the agribusiness community in stopping the UFW, setting the stage for 11 subsequent years of conflict. Underlying its success was the UFW's commitment to its mission and its growing capacity to accomplish it more effectively. As Chavez often said, "It's not so important that you make the right decision. What is important is that you learn to do all you can to make the decision you do make the right decision" (Ganz 1993).

Although the Schenley Boycott was a crucial turning point on NFWA's path to success, its significance for this analysis is not as an isolated event—or "good tactic"—but rather as an outcome of the strategic capacity the NFWA had generated since the beginning of the grape strike. NFWA's decision to boycott Schenley, like the other choices examined in this study, was a strategic response to a new challenge—how to maintain pressure on the employers despite the end of the grape harvest. Just as the effectiveness of NFWA's responses were related to the breadth of information to which it had access, its heuristic processes, and the depth of its motivation, the ineffectiveness of AWOC's responses were due to the limited salience of its information, the constraints on its creativity, and a

marginal level of commitment. As Green said when asked to support NFWA's tactics: "This is an honest to goodness trade union fight, not a civil rights demonstration. . . . I am relying on union support. . . . The NFWA is administered by ministers. . . . We (AWOC) will continue in our own union way" (Taylor 1975, p. 155).

CONCLUSIONS

This article began by posing the question of why organizing success came to the fledgling UFW and not to the well-established union with which it found itself in competition. Studying the influence of strategy reveals the role of resourcefulness in power—as mythically memorialized in tales of David and Goliath or Odysseus and the Trojans. One way groups compensate for a lack of material resources is through creative strategy, a function of access to a diversity of salient information, heuristic facility, and motivation—a result of the way the composition of leadership teams and organizational structures influence interaction with the environment. Changing environments generate new opportunities—and constraints—but the significance of those opportunities or constraints emerges from the hearts, heads, and hands of the actors who develop the means of acting upon them.

As summarized in table 2, the source of difference in the strategic capacity of the two groups was in observable differences in leadership and organization. The contrast in the biography, networks, and repertoires of the leadership and deliberative processes, resource flows, and accountability structures of the organizations could not have been greater. Because strategy unfolds as a process, it is important to pay attention to the mechanisms that generate it—not only to the role of specific strategies in specific outcomes. Since strategy is interactive, getting it "right" in a big way is likely to be evidence of having learned how to "get it right" in numerous small ways—and doing it time after time. This can only be studied by observing organizations over time.

The UFW's strategy thus turned out to be more effective than that of the AWOC because of the way in which it was developed. It drew on elements of an ethnic labor association (reminiscent of earlier organizing attempts by farmworkers of color), a union, and community organizing drives in a new synthesis that went far beyond its individual components as its founders engaged environmental challenges by adapting familiar repertoires to new uses. The UFW's response to the crisis precipitated by the grape strike was to draw on the Civil Rights movement to reframe its effort as a farmworker movement. This then led to development of a "dual strategy" based on mobilization of workers (without whom there would have been no people, no cause, and no movement) along with the

TABLE 2

COMPARISON OF AWOC AND UFW STRATEGIC CAPACITY

	AWOC	UFW
Leadership:		
Biography	Little diversity of experience No salient local knowledge Professional commitment	Diversity of experience Salient local knowledge, broader context Personal, vocational commitment
Networks	No strong ties to constituency Few salient weak ties Little diversity of ties	Strong ties to constituencies Weak ties across constituencies Diversity of ties
Repertoires	No salience to constituencies Little diversity of repertoires	Salience to constituencies Diversity of repertoires
Organization:		
Deliberation	No regular meetings No strategy sessions Closed to diverse perspectives Not authoritative	Regular meetings Regular strategy sessions Open to diverse perspectives Authoritative
Resource flows	Resources flow top down Resources flow from outside: single source Resources depend on internal politics Based on financial resources Little strategic autonomy	Resources flow bottom up Resource flow inside and outside: multiple sources Resources depend on task effectiveness Based on people resources Strategic autonomy
Accountability	No constituency accountability Hierarchical accountability Bureaucratic leadership selection	Constituency accountability Mutual accountability Democratic, entrepreneurial selection

mobilization of urban supporters (without whom there would have been no financial, political, and economic resources). By recontextualizing the arena of combat to reach beyond the fields to the cities, the UFW turned the moral tables on the growers, exposing what growers considered to be a legitimate exercise of their authority as illegitimate in the public domain (much as had occurred in the Civil Rights movement). The significance of this strategic stream for this article is not in its particulars—although it points to important lessons about targeting, timing, and tactics—but in that it emerged from a strategic capacity that could have generated a different strategic stream in different circumstances.

Focusing on strategic capacity—or its absence—may also help account for the outcome of other organizing efforts. In the course of American labor history, the question remains of why the breakaway CIO successfully organized industrial workers in the 1930s, a job the AFL would not or could not do. A more recent question, however, is why the semi-independent Harvard Union of Clerical and Technical Workers (HUCTW) succeeded where the United Auto Workers (UAW) District 65 repeatedly failed (Hoerr 1997). Evidence suggests the UAW was hamstrung by factors similar to those that limited the AWOC, while the HUCTW's effectiveness was rooted in strategic capacity very similar to that of the UFW—although it played out in different specific strategies. Similarly, scholars have only begun to evaluate the efforts of a revived AFL-CIO to organize today (Bronfenbrenner and Juravich 1998), but most of the attention has been on tactical efficacy rather than on the role of leadership and organization in building strategic capacity.

Understanding strategic capacity may be useful not only in unraveling social movements—it can help explain the outcomes of other conflicts in which new groups are far more effective than well-established ones, such as the high-tech industry. Although Stinchcombe (1965) argued new organizations must overcome a “liability of newness,” my research shows there may be a “liability of senescence” and that newness can be an asset. AWOC leaders, selected for reasons that had little to do with the needs of the environment within which they were to work, developed strategy within an organizational setting better equipped to reproduce past routines than to innovate new ones. Ironically, the abundance of internal resources to which well-established groups have access may make it harder to innovate by making it easier for them to keep doing the same thing wrong. New groups, on the other hand, often lack conventional resources, but the richness of their strategic capacity—aspects of their leadership and organization in relation to the environment specified here—can offset this.

Finally, this approach offers fresh ways to make intractable problems actionable by holding out the possibility of change. People can generate

the power to resolve grievances not only if those with power decide to use it on their behalf, but also if they can develop the capacity to out think and outlast their opponents—a matter of leadership and organization. As students of “street smarts” have long understood, “resourcefulness” can sometimes compensate for a lack of resources. While learning about how the environment influences actors is very important, learning more about how actors influence the environment is the first step not only to understanding the world, but to changing it.

APPENDIX

Biographical Sketches

AWOC Leadership

George Meany, 66, white Irish American, Roman Catholic, married, three children, high school graduate, a New Yorker, was president of the AFL-CIO.¹⁸ A lifelong leader of the plumbers' union, the building trades, and the AFL—as was his father before him—he had no real interest in organizing farmworkers. He once asked, “Why should we worry about organizing groups of people who do not appear to want to be organized?” (Zieger 1987, p. 342). A strong anticommunist, he was suspicious of the emergent Civil Rights movement, denying AFL-CIO support for the 1963 March on Washington. He possessed an impressive tactical repertoire in internal union politics and legislative lobbying, which, combined with a lack of organizing experience, gave him a marked preference for legislative strategy (Mitchell 1979; London and Anderson 1970; Fink 1974; Zieger 1987).

John Livingston, 52, white, Protestant, married, was the director of organizing. Growing up on a Missouri farm, Livingston attended high school for two years before going to work for General Motors in St. Louis. He learned his tactical repertoire as a leader in the organization of the auto industry in the 1930s. Originally recruited by Norman Smith, he rose to become a vice president of the UAW, a position he left to join the AFL-CIO in 1955. After his retirement in 1965, he became director of union relations for the National Alliance of Businessmen. Although he had no networks linking him to the world of farmworkers, they did extend to former UAW organizers, which is how he found Norman Smith (Fink 1974; Reuther 1976; Mitchell 1979; London and Anderson 1970; Meister and Loftis 1977; Zieger 1987; Lichtenstein 1968).

Norman Smith, 62, white, Protestant, unmarried, also from a Missouri

¹⁸ Ages given are at time participants undertook farmworker organizing.

farm family, was the first AWOC director. A UAW industrial organizer in the 1930s who had recruited Livingston. After serving as a Seabee in World War II, Smith had been an industrial supervisor for 18 years. He had no agricultural experience, spoke no Spanish, and had no links to the farmworker world. The tactical repertoire he learned in the auto industry influenced his search for an equivalent of the "plant gate," the early morning shape-ups that led him to target a segment of the work force least likely to provide a stable membership base (Mitchell 1979; London and Anderson 1970; Meister and Loftis 1977; Jenkins 1985; Anderson 1996).

A. C. Green, 60, white, Protestant, was the second AWOC director. Green had been a Plasterers Union official since his youth, and for the previous 12 years, director of the California AFL-CIO COPE (Committee on Political Education). Despite growing up in the San Joaquin Valley town of Modesto, a center for Anglo cannery workers, he had no links to the farmworker world. The tactical repertoire he learned in the building trades was the source of the labor contractor tactics that proved so inappropriate in this setting (London and Anderson 1970; Meister and Loftis 1977; Jenkins 1985; Anderson 1996).

AWOC Staff

Dr. Ernesto Galarza, 54, Mexican, Roman Catholic, married, three children, was hired by Livingston to assist Smith. Galarza, who emigrated with his family from Mexico in 1910, had grown up in Sacramento, won scholarships to Occidental College, Stanford, and earned a Ph.D. in economics at Columbia. From 1934 to 1946 he did research for the Pan American Union, becoming an expert on the wartime bracero program. In 1947, hired by H. L. Mitchell as research director for the AFL's National Farm Labor Union (NFLU), he settled with his family in San Jose. After leading a number of unsuccessful organizing attempts in the early 1950s, he dedicated himself to "exposure" of the abuses of the bracero program, publishing the influential *Strangers in Our Fields* in 1956. Passed over for the AWOC directorship, he resigned after six months in a dispute over whether the old NFLU (now called the National Agricultural Workers Union [NAWU]) would have jurisdiction to workers organized by the AWOC (London and Anderson 1970; Anderson 1996).

Father Tom McCullough, 37, white, Roman Catholic, son of a Bay Area trade unionist, was one of a network of Roman Catholic priests with a ministry to farmworkers, the California Mission Band. Meeting as seminarians at St. Patrick's Seminary, Menlo Park, they were constituted as the Missionary Apostolate of the San Francisco Diocese in 1950. McCullough's associate was Father Donald McDonnell, the San Jose priest who

interested Cesar Chavez in organizing. In the late 1950s, McCullough organized a mutual benefit association among his Stockton farmworker parishioners, which was a model for the UFW's work a number of years later. He merged his group into the AWOC when it began, thinking "they were the professionals" (London and Anderson 1970; Meister and Loftis 1977; Jenkins 1985; Anderson 1996).

Dolores Huerta, 30, Mexican American, Roman Catholic, possessed years of community organizing experience working with Fred Ross and Cesar Chavez. She would later become Chavez's second in command (see below for more).

Henry Anderson, 32, white, married, had grown up in the Bay Area, served in World War II, was educated at Pomona College and Stanford, and earned masters degrees in sociology and public health at the University of California, Berkeley. Doing research for the California Department of Public Health, he became an expert on the abuses of the bracero program and met Father McCullough and others involved in early farmworker organizing. Hired as AWOC research director by Smith in 1959, he led the volunteer organizing effort during the AWOC "hiatus" and authored a 1961 organizing plan reflecting McCullough's work and which presaged the approach Chavez would take (London and Anderson 1970; Meister and Loftis 1977; Jenkins 1985; Anderson 1996).

Larry Itliong, 47, Filipino, Roman Catholic, married, immigrated from the Philippines at 15, worked in the fields, organized for UCAPAWA CIO, was leader of ILWU local 7, formed his own Farm Labor Union in 1956, and was president of Filipino Community of Stockton, 1959. Hired as an organizer by AWOC in 1960, he led the Delano Grape Strike in 1956. He became assistant to Chavez in the merger of UFWOC (Scharlin and Villanueva 1992; Taylor 1975; Watson 1999).

UFW Leadership

Cesar Chavez, president of the NFWA, was 35 in 1962. He was Mexican American, Roman Catholic, and married with eight children. He grew up in an immigrant Mexican family who, when he was 10, lost a small Arizona farm in the depression and became migrants. After he served in the Navy, he and his wife Helen settled in San Jose, where he found work in a lumberyard. He became active in the church, and Father Donald McDonnell, his parish priest, got him interested in organizing. McDonnell was associated with McCullough in the California Mission Band. He began learning organizing in 1952 when recruited by Fred Ross, an associate of Saul Alinsky's, to build the first statewide Mexican American civic association in California, the CSO. In this work, he developed statewide networks of Mexican American activists, religious groups, liberal Demo-

crats, and unions—which went far beyond ties he retained within the farmworker community. His first union organizing experience was a 1958 United Packing House Workers drive in the Oxnard citrus packing sheds, an effort backed by the CSO. Early in 1962, during the hiatus in the AWOC, he resigned as executive director of the CSO when it rejected his proposal to organize farmworkers. Relying on his savings, he moved to Delano, where he had family, to begin organizing. He also turned down a Kennedy administration job offer to be Peace Corps director in Venezuela (Levy 1975; Taylor 1975; Meister and Loftis 1977; Ross 1989; Daniel 1987).

Dolores Huerta, secretary of the NFWA, was 30 in 1959 when she became active with the AWOC, and 33 when she joined Chavez in 1962. She was Mexican American, Roman Catholic, and married with six children. Huerta was a native of New Mexico where her father had organized for the miners' union and served for a term in the New Mexico legislature. She grew up in Stockton where her mother ran a boarding house. After graduating from the College of the Pacific, she was recruited by Ross for work with the CSO in 1953. After working briefly with the AWOC in 1959, she led a successful CSO campaign to extend California old-age pensions to noncitizens. She joined Chavez part time in 1962, going to work full time with the NFWA in 1964 (Levy 1975; Taylor 1975; Meister and Loftis 1977).

Antonio Orendain, treasurer of the NFWA, was a farmworker, about 30, immigrated from Mexico in 1956, Roman Catholic, and married. He and his wife Raquel had been active leaders in the CSO since 1958 (Brown 1972; Levy 1975; Padilla 1999).

Gilbert Padilla, NFWA vice president, was 35, Mexican American, Roman Catholic, and married with four children. Padilla had grown up on a farm outside Fresno, served in the Army, worked as a cleaner, and had been recruited by Chavez into the CSO in 1956, first as a leader and later as a full-time organizer (Brown 1972; Levy 1975; Taylor 1975; Smith 1987; Drake 1997; Padilla 1999).

Julio Hernandez, NFWA vice president, was 41, immigrated from Mexico in 1944, Roman Catholic, and married with nine children. He and his wife, Josefina, were farmworkers, had been labor contractors, joined CSO, and were among the first recruited by Chavez for the NFWA. Hernandez served as president of the Farm Workers Credit Union (Brown 1972; Levy 1975; Taylor 1975; Smith 1987).

Chris Hartmire, director of the California Migrant Ministry (CMM), was 29, white, Presbyterian, and married with two children. After graduating from Princeton and serving in the Navy for three years, he attended Union Theological Seminary (UTS). There, in 1954, he began work in the East Harlem Ministry and was recruited by Doug Stills, the Alinsky-

trained director of the CMM, an agency of the National Council of Churches. The CMM was in transition from a social service program to a social action program. He met Chavez and Ross in 1959 when he was sent to "train" with them in the CSO. Hartmire had become director of the CMM in 1961, the year he met Alinsky and participated in a CORE-sponsored "Freedom Ride" (Levy 1975; Taylor 1975; Meister and Loftis 1977; Smith 1987; Drake 1997; Hartmire 1996).

Jim Drake, a California Migrant Ministry "minister-organizer" was 24, white, United Church of Christ, and married with two children. Drake, who also attended Union Theological Seminary, was one of Hartmire's first recruits for his new CMM program. He grew up in a Mexican farmworker community in Southern California where his father had served as a teacher and minister (Levy 1975; Taylor 1975; Meister and Loftis 1977; Hartmire 1996; Drake 1997).

UFW Volunteers

New young farmworker leaders emerged from striking farmworker families. They served as picket captains, organizers, and boycott directors; some later rose to top leadership positions in the union. Eliseo Medina, for example, an 18-year-old member of a striking family, and a Mexican immigrant, was trained as an organizer, became Chicago boycott director, director of the UFW's field offices, and was later elected as vice president of the UFW. Marcos Munoz, a 20-year-old Mexican immigrant from Bakersfield, who was a very talented organizer but could neither read nor write English or Spanish, became the Boston boycott director. Maria Saluado, 22, and her sisters, Petra and Antonia, grew up in a farmworker family, began working in the fields as children, and became boycott organizers in Philadelphia, New York, Chicago, and elsewhere (Rose 1995). Of 40 boycott cities operating at the peak of the grape boycott, some 35 were led by new farmworker leaders (Brown 1972).

Students involved in the Civil Rights movement came to Delano as volunteers, assuming a wide variety of roles as organizers, boycotters, administrators, and so on (Nelson 1966; Taylor 1975; Levy 1975). Luis Valdez, for example, a 23-year-old Chicano student from a farmworker family, came to Delano and organized the Teatro Campesino, a strikers' theater troop that generated songs and skits for the strike and served as a model for urban Chicano theater throughout the Southwest (Mathiessen 1969). Jessica Govea, an 18-year-old Chicana from Bakersfield whose family had been active in the CSO, dropped out of college, went to work in the union's service program, became an organizer, designed the union's medical program, and was elected to the national executive board. After growing up in Bakersfield, attending Harvard for three years, and serving

with SNCC for two years, I came to the UFW at age 22, served in a variety of roles, eventually became director of organizing, and was elected to the national executive board (Levy 1975).

Religious activists inspired by the Civil Rights movement, recruited by the Migrant Ministry, or motivated by Vatican II came to serve in a variety of roles (Hartmire 1996; Chatfield 1996; Drake 1997). Leroy Chatfield, for example, a former Christian Brother who had been associated with the Catholic Worker, became director of the union's service center and medical fund.

Lawyers drawn to public service, but attracted to the opportunity to practice more "political" law than was possible through legal services corporations, came to serve as a full-time and part-time resource for the UFW (Chatfield 1996; Cohen 1995; Levy 1975). Jerry Cohen, for example, resigned from California Rural Legal Assistance (CRLA) to come to work full time for the UFW, eventually becoming its general counsel and authoring the California Agricultural Labor Relations Act of 1975.

TABLE A1
LEADERSHIP COMPARISON

Name	Age	Race/Ethnicity	Religion
AWOC leadership:			
George Meany	66	White/Irish-American	Roman Catholic
John Livingston	52	White	Protestant
Norman Smith	61	White	Protestant
A. C. Green	60	White	Protestant
AWOC staff:			
Father Tom McCullough	37	White	Roman Catholic
Dolores Huerta	30	Mexican-American	Roman Catholic
Henry Anderson	32	White	...
Larry Itliong	50	Filipino	Roman Catholic
Dr. Ernesto Galarza	54	Mexican	Roman Catholic
NFWA leadership:			
Cesar Chavez	35	Mexican-American	Roman Catholic
Dolores Huerta	33	Mexican-American	Roman Catholic
Antonio Orendain	30	Mexican	Roman Catholic
Gilbert Padilla	35	Mexican-American	Roman Catholic
Julio Hernandez	41	Mexican	Roman Catholic
Rodrigo Terronez	35	Mexican	Roman Catholic
Chris Hartmire	29	White	Presbyterian
Jim Drake	24	White	United Church of Christ
NFWA volunteers:			
Eliseo Medina	18	Mexican	Roman Catholic
Marcos Munoz	23	Mexican	Roman Catholic
Maria Salgado	22	Mexican	Roman Catholic
Luis Valdez	23	Mexican-American	Roman Catholic
Marshall Ganz	22	White	Jewish
Jessica Govea	18	Mexican-American	Roman Catholic
Leroy Chatfield	27	White	Roman Catholic
Jerry Cohen	27	White	Jewish

SOURCE.—Anderson (1996); Brown (1972); Chatfield (1996); Cohen (1995); Daniel (1987); Drake (1997); Fink (1974); Hartmire (1996); Jenkins (1985); Levy (1975); London and Anderson (1970); Mathiessen (1969); Medina (1998); Meister and Loftis (1977); Mitchell (1979); Nelson (1966); Padilla (1999); Reuther (1976); Rose (1995); Ross (1989); Scharlin and Villanueva (1992); Smith (1987); Taylor (1975); Zieger (1987).

NOTE.—Age = time at which participants undertook farmworker organizing. Huerta is listed twice—she was 30 when she began working with AWOC and 33 when she began working with the UFW.

Regional Background	Position	Family Background
New York, District of Columbia	AFL-CIO president	Construction union officials
St. Louis, Detroit, District of Columbia	Director of organizing	Missouri farm family
Detroit, Los Angeles	AWOC director	
Modesto	AWOC director	
Oakland, Stockton	Volunteer organizer	Building trades, union
New Mexico, Stockton	Organizer	Mining, boarding house
Palo Alto, Berkeley	AWOC research director	Professional, teaching
Philippines, Stockton, Delano	Organizer	Immigrant, rural Philippines
Sacramento, New York, District of Columbia, San Jose	AWOC assistant director	Immigrant farmworkers
Yuma, San Jose, Delano	NFWA president	Immigrant small farmers, farmworkers
New Mexico, Stockton	NFWA secretary	Mining, small business
Mexico, Hanford	NFWA treasurer	Immigrant farmworkers
Los Banos, Hanford	NFWA vice president	Labor, immigrant farmworkers
Mexico, Corcoran	NFWA vice president	Labor, immigrant
Mexico, Hanford	NFWA vice president	Immigrant farmworkers
New York, Los Angeles	CMM director	Professional, business
Oklahoma, Indio, New York	CMM minister organizer	Professional, ministry, teaching
Mexico, Delano	Organizer/boycotter	Immigrant farmworkers
Mexico, Bakersfield	Organizer/boycotter	Immigrant farmworkers
Mexico, Delano	Organizer/boycotter	Immigrant farmworkers
Delano, San Jose	Teatro Campesino	Farmworkers, laborers
Bakersfield	Organizer	Professional, ministry, teaching
Bakersfield	Organizer, social services	Immigrant railroad, farmworkers
Sacramento	Organizer, social services	Farmers, business
District of Columbia, Japan	Lawyer	Professional, medical, military

Abbreviations for organizations are AFL (American Federation of Labor); CIO (Congress of Industrial Organizations); CMM (California Migrant Ministry); CRLA (California Rural Legal Assistance); CSO (Community Service Organization); ILWU (International Longshoremen's and Warehousemen's Union); SNCC (Student Nonviolent Coordinating Committee); UAW (United Auto Workers); UTS (Union Theological Seminary). Other abbreviations are M (married); A (active spouse); S (single); D (divorced); and C (number of children).

TABLE A1 (Continued)

Name	Family	Education, Work, Organizing Experience
AWOC leadership:		
George Meany	M, 3C	High school, plumber, AFL union official
John Livingston	M	10th grade, autoworker, UAW union official
Norman Smith	S	Autoworker, UAW organizer, Seabee industrial supervisor
A. C. Green	S	AFL union official, California COPE director
AWOC staff:		
Father Tom McCullough	S	Franciscan priest, farmworker organizer, AWO
Dolores Huerta	M	College of Pacific, CSO organizer, lobbyist
Henry Anderson	M	UC Berkeley, M.D., researcher, advocate
Larry Itliong	M	6th grade, farmworker, UCAPAWA organizer, community leader
Dr. Ernesto Galarza	M, 2C	Occidental, Stanford, Columbia Ph.D., Pan-American Union, researcher, advocate, NFLU, NAWU
NFWA leadership:		
Cesar Chavez	M, A, 8C	8th grade, farm work, Navy, laborer, CSO, organizer
Dolores Huerta	D, 6C	College of Pacific, CSO organizer, lobbyist, AWOC
Antonio Orendain	M, A, 2C	Farm work, CSO
Gilbert Padilla	M, 4C	Farm work, Army, small business, CSO
Julio Hernandez	M, A, 9C	Farm work, labor contractor, CSO
Rodrigo Terronez	M	Farmworker, CSO
Chris Hartmire	M, 2C	Princeton, UTS, East Harlem Ministry, Navy
Jim Drake	M, 2C	UTS
NFWA volunteers:		
Eliseo Medina	S	Farm work
Marcos Munoz	M, 1C	Farm work
Maria Saluado	S	Farm work
Luis Valdez	S	San Jose State, SNCC, San Francisco mime troop
Marshall Ganz	S	Harvard, SNCC
Jessica Govea	S	Bakersfield College, CSO
Leroy Chatfield	M	Christian Brothers, vice principal, teaching
Jerry Cohen	M, 1C	Amherst College, UC Berkeley Law School, CRLA

Work Commitment	Network Affiliation	Repertoires
Professional	AFL-CIO, Democrats, Roman Catholics	Union politics, legislative lobbying, elite politics
Professional	UAW, Democrats	CIO 1930-40s auto organizing, union politics
Professional	UAW, Democrats	CIO 1930s auto organizing, supervisor
Professional	Building trades, Democrats	Building trades, COPE
Vocational	Farmworkers, Roman Catholics, farmworker advocates	Religious community, mutual benefit association
Personal/vocational	CSO, Democrats, farmworkers	Community organizing, lobbying
Vocational	Students, liberals, farmworker advocates	Research, advocacy
Personal/vocational	Filipino crew leaders, union organizers	Cannery organizing, crew and community leadership
Personal/vocational	NFLU, farmworker advocates, Democrats	Research, lobbying, and advocacy
Personal/vocational	Family, farmworkers, CSO, Roman Catholics, Democrats, Mexican Americans	Community organizing, political organizing
Personal/vocational	Family, CSO, Democrats, liberals	Community organizing, lobbying
Personal/vocational	Family, farmworkers, CSO	Community organizing
Personal/vocational	Family, farmworkers, CSO	Community organizing
Personal	Family, farmworkers	Crew leader, community organizing
Personal	Family, farmworkers, CSO	Community organizing
Vocational	CMM, California and U.S. church leaders, farmworker advocates	Seminary, community organizing
Personal	CMM	Seminary, community organizing
Personal	Family, farmworkers	
Personal	Family, farmworkers	
Personal	Family, farmworkers	
Personal/vocational	Family, student activists, liberals	Student organizing, political theater
Vocational	Civil rights groups, student groups	Civil rights organizing
Personal/vocational	Family, CSO	CSO
Vocational	Roman Catholic	Religious community, teaching
Vocational	Legal activists, lawyers	Advocacy, antiwar protests

REFERENCES

- Abelson, Robert P. 1981. "Psychological Status of the Script Concept." *American Psychologist* 36:715-29.
- Aldrich, H. E., and E. R. Auster. 1986. "Even Dwarfs Started Small: Liabilities of Age and Size and Their Strategic Implications." Pp. 165-98 in *Research in Organizational Behavior*, vol. 8. Edited by B. M. Staw and L. L. Cummings. Greenwich, Conn.: JAI.
- Alexander, Victoria D. 1998. "Environmental Constraints and Organizational Strategies: Complexity, Conflict and Coping in the Nonprofit Sector." Pp. 272-90 in *Private Action and the Public Good*, edited by Walter W. Powell and Elisabeth Clemens. New Haven, Conn.: Yale University Press.
- Alinsky, Saul. 1971. *Rules for Radicals*. New York: Vintage.
- Amabile, Theresa M. 1988. "Creativity and Innovation in Organizations." Pp. 123-67 in *Research in Organizational Behavior*, vol. 10. Edited by B. M. Staw and L. L. Cummings. Greenwich, Conn.: JAI.
- . 1996. "A Theoretical Framework." Pp. 81-130 in *Creativity in Context*. Boulder, Colo.: Westview Press.
- Anderson, Henry. 1961. *To Build a Union: Comments on the Workers. Organization of Agricultural Workers*. Stockton, Calif.: Self-published.
- Banaszak, Lee Ann. 1996. *Why Movements Succeed or Fail: Opportunity, Culture and the Struggle for Woman Suffrage*. In *Princeton Studies in American Politics*, edited by Ira Katznelson and Theda Skocpol. Princeton, N.J.: Princeton University Press.
- Bandura, A. 1989. "Human Agency in Social Cognitive Theory." *American Psychologist* 44 (9): 1175-84.
- Barnard, John. 1983. *Walter Reuther and the Rise of the Auto Workers*. Boston: Little, Brown.
- Bergman, J. 1979. "Energy Levels: An Important Factor in Identifying and Facilitating the Development of Giftedness in Young Children." *Creative Child and Adult Quarterly* 4:181-88.
- Bernstein, B. 1975. "Social Class, Language and Socialization." Pp. 170-89 in *Class Codes and Control: Theoretical Studies towards a Sociology of Language*, 2d ed. New York: Schocken Books.
- Boone, C., W. van Olfen, A. van Witteloostuijn. 1998. "Psychological Team Make-up as a Determinant of Economic Firm Performance—An Experimental Study." *Journal of Economic Psychology* 19 (1): 43-73.
- Bourdieu, Pierre. 1990. *The Logic of Practice*. Stanford, Calif.: Stanford University Press.
- Bower, J. L. 1970. *Managing the Resource Allocation Process: A Study of Corporate Planning and Investment*. Homewood, Ill.: Richard D. Irwin.
- Boycott Newsletter*. 1965. Ganz papers.
- Bronfenbrenner, Kate, and Tom Juravich. 1998. "It Takes More Than House Calls: Organizing to Win with a Comprehensive Union-Building Strategy." Pp. 19-36 in *Organizing to Win: New Research on Union Strategies*, edited by Kate Bronfenbrenner, Sheldon Friedman, Richard W. Hurd, Rudolph A. Oswald, and Ronald L. Seeber. Ithaca, N.Y.: ILR Press.
- Brown, Jerald. 1972. *The United Farm Workers Grape Strike and Boycott, 1965-70: An Evaluation of the Culture of Poverty Theory*. Doctoral dissertation. Cornell University, Department of Anthropology.
- Brown, Shona L., and Kathleen M. Eisenhardt. 1997. "The Art of Continuous Change: Linking Complexity Theory and Time-Paced Evolution in Relentlessly Shifting Organizations." *Administrative Science Quarterly* 42 (1): 34-56.
- . 1998. *Competing on the Edge: Strategy as Structured Chaos*. Boston: Harvard Business School Press.

- Bruner, Jerome. 1990. *Acts of Meaning*. Cambridge, Mass.: Harvard University Press.
- Burgelman, R. A. 1991. "Intraorganizational Ecology of Strategy Making and Organizational Adaptation: Theory and Field Research." *Organization Science* 2 (3): 239-62.
- California Assembly Committee on Agriculture. 1969. *The California Farm Workforce: A Profile*. Sacramento, Calif.: California Advisory Committee on Farm Labor Research.
- Campbell, Donald T. 1960. "Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes." *Psychological Review* 67:380-400.
- . 1975. "Degrees of Freedom and the Case Study." *Comparative Political Studies* 8 (2): 178-93.
- Campbell, John L. 1997. "Mechanisms of Evolutionary Change in Economic Governance: Interaction, Interpretation, and Bricolage." Pp. 10-32 in *Evolutionary Economics and Path Dependence*, edited by Lars Magnusson and Jan Ottosson. Cheltenham: Edward Elgar.
- Carmines, Edward G., and James A. Stimson. 1989. *Issue Evolution: Race and the Transformation of American Politics*. Princeton, N.J.: Princeton University Press.
- Chambers, J. A. 1964. "Relating Personality and Biographical Factors to Scientific Creativity." *Psychological Monographs* 78 (7): whole no. 584.
- Chandler, Alfred D. 1962. *Strategy and Structure: Chapters in the History of the American Industrial Enterprise*. Cambridge, Mass.: MIT Press.
- . 1977. *The Visible Hand: The Managerial Revolution in American Business*. Cambridge, Mass.: Belknap Press.
- Child, John. 1972. "Organizational Structure, Environment and Performance: The Role of Strategic Choice." *Sociology* 6 (1): 1-22.
- Chong, Dennis. 1991. *Collective Action and the Civil Rights Movement*. Chicago: University of Chicago Press.
- Clausewitz, Carl Von. 1832. "Book Three, of Strategy in General." Pp. 241-98 in *On War*, edited by Anatol Rapoport. London: Penguin Books.
- Clemens, Elisabeth. 1996. "Organizational Form as Frame: Collective Identity and Political Strategy in the Labor Movement, 1880-1920." Pp. 205-26 in *Comparative Perspectives on Social Movements*, edited by Doug McAdam, John McCarthy, and Meyer Zald. New York: Cambridge University Press.
- Cohen, Michael D., James G. March, and Johan P. Olsen. 1972. "Garbage Can Model of Organizational Choice." *Administrative Science Quarterly* 17:1-25.
- Coleman, James. 1990. *Foundations of Social Theory*. Cambridge, Mass.: Harvard University Press.
- Conti, R., T. M. Amabile, and S. Pollak. 1995. "Problem Solving among Computer Science Students: The Effects of Skill, Evaluation Expectation and Personality on Solution Quality." Paper presented at the Annual Meeting of the Eastern Psychological Association, Boston, April.
- Craig, R. B. 1970. *Interest Groups and the Foreign Policy Process: A Case Study*. Doctoral dissertation. University of Missouri, Department of Political Science.
- Cropley, A. 1990. "Creativity and Mental Health in Everyday Life." *Creativity Research Journal* 3:167-78.
- Damanpour, Fariborz. 1991. "Organizational Innovation: A Meta-Analysis of Effects of Determinants and Moderators." *Academy of Management Journal* 34 (3): 555-90.
- D'Andrade, Roy G. 1992. "Schemas and Motivation." Pp. 23-44 in *Human Motives and Cultural Models*, edited by Roy D'Andrade and Claudia Strauss. New York: Cambridge University Press.
- Daniel, Cletus E. 1981. *Bitter Harvest: A History of California Farm Workers, 1870-1941*. Ithaca, N.Y.: Cornell University Press.
- . 1987. "Cesar Chavez and the Unionization of California Farm Workers." Pp.

- 350–82 in *Labor Leaders in America*, edited by Melvyn Dubovsky and Warren Van Tine. Urbana: University of Illinois Press.
- Deci, E. L., and R. M. Ryan. 1980. "The Empirical Exploration of Intrinsic Motivational Processes." Pp. 39–80 in *Advances in Experimental Social Psychology*, edited by L. Berkowitz. New York: Academic Press.
- DiMaggio, Paul. 1997. "Culture and Cognition." *Annual Review of Sociology* 23:263–87.
- Douglas, Mary. 1986. *How Institutions Think*. Syracuse, N.Y.: Syracuse University Press.
- Duncan, R. B. 1973. "Multiple Decision Making Structures in Adapting to Environmental Uncertainty: The Impact on Organizational Effectiveness." *Human Relations* 26 (3): 273–92.
- Dunne, John Gregory. 1967. *Delano, the Story of the California Grape Strike*. New York: Farrar, Straus & Giroux.
- Durkheim, Émile. 1915. *Elementary Forms of Religious Life*. New York: Macmillan.
- Eisenhardt, Kathleen M., and Shona L. Brown. 1998. "Time Pacing: Competing in Markets That Won't Stand Still." *Harvard Business Review* 76 (2): 59–69.
- Eisenhardt, Kathy M., and Claudia Bird Schoonhoven. 1990. "Organizational Growth: Linking Founding Team, Strategy, Environment, and Growth among U.S. Semiconductor Ventures, 1978–1988." *Administrative Science Quarterly* 35:504–29.
- Emerson, Richard. 1962. "Power-Dependence Relations." *American Sociological Review* 27:31–44.
- Emirbayer, M., and J. Goodwin. 1994. "Network Analysis, Culture, and the Problem of Agency." *American Journal of Sociology* 99:1411–54.
- Fink, Gary M. 1974. *Biographical Dictionary of American Labor Leaders*. Westport, Conn.: Greenwood Press.
- Fiske, Susan, and Shelly E. Taylor. 1991. *Social Cognition*. New York: McGraw-Hill.
- Freeman, Jo. 1979. "Resource Mobilization and Strategy: A Model for Analyzing Social Movement Organizations." Pp. 167–89 in *The Dynamics of Social Movements: Resources Mobilization, Social Control, and Tactics*, edited by M. N. Zald and J. D. McCarthy. Cambridge, Mass.: Winthrop.
- Freeman, John, Glenn R. Carroll, and Michael T. Hannan. 1983. "The Liability of Newness: Age Dependence in Organizational Death Rates." *American Sociological Review* 48 (5): 692–710.
- Gamson, William. 1975. *The Strategy of Social Protest*. Belmont, Calif.: Wadsworth Publishing.
- . 1990. "Challenging Groups since 1945." Pp. 145–80 in *Strategy of Social Protest*. Belmont, Calif.: Wadsworth Publishing.
- . 1991. "Commitment and Agency in Social Movements." *Sociological Forum* 6 (1): 27–50.
- . 1992. *Talking Politics*. New York: Cambridge University Press.
- Gamson, William A., Bruce Fireman, and Steven Rytina. 1982. *Encounters with Unjust Authority*. Homewood, Ill.: Dorsey Press.
- Gamson, William A., and David S. Meyer. 1996. "Framing Political Opportunity." Pp. 275–90 in *Comparative Perspectives on Social Movements*, edited by Doug McAdam, John D. McCarthy, and Mayer N. Zald. Cambridge: Cambridge University Press.
- Ganz, Marshall. 1993. Lecture notes on Cesar Chavez. Harvard University, Kennedy School of Government. Unpublished.
- . 1994. Notes on the March to Sacramento. Unpublished.
- Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. Cambridge: Prentice-Hall.
- Gentner, Dedre. 1989. "Mechanisms of Analogical Learning." Pp. 199–239 in *Similar-*

- ity and Analogical Reasoning, edited by S. Vosiadou and A. Ortony. Cambridge: Cambridge University Press.
- Getzels, J., and M. Csikszentmihalyi. 1976. *The Creative Vision: A Longitudinal Study of Problem Finding in Art*. New York: Wiley-Interscience.
- Glaser, Barney G., and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Hawthorne, N.Y.: Aldine de Gruyter.
- Glover, J. A., and F. Sautter. 1977. "Relation of Four Components of Creativity to Risk-Taking Preferences." *Psychological Reports* 41:227-30.
- Goffman, Erving. 1974. *Frame Analysis*. New York: Harper & Row.
- Gorman, Patrick. 1959. "The Forgotten Farm Worker." Pp. 31-37 in *IUD Digest*. Washington, D.C.: Industrial Union Department, AFL-CIO.
- Granovetter, Mark. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78:1360-80.
- Griswold, Wendy. 1987. "The Fabrication of Meaning: Literary Interpretation in the United States, Great Britain, and the West Indies." *American Journal of Sociology* 92:1077-117.
- Hackman, J. R. 1990. "Creating More Effective Work Groups in Organizations." Pp. 479-504 in *Groups That Work*, edited by J. R. Hackman. San Francisco: Jossey-Bass.
- Hackman, Richard, and Richard Walton. 1986. "Leading Groups in Organizations." Pp. 72-119 in *Designing Effective Work Groups*, edited by Paul Goodman. San Francisco: Jossey-Bass.
- Hamel, Gary. 1996. "Strategy as Revolution." *Harvard Business Review* 74 (69): 69-78.
- Hamel, Gary, and C. K. Prahalad. 1989. "Strategic Intent." *Harvard Business Review* 67 (3): 63-77.
- Hannan, Michael T., John Freeman. 1977. "The Population Ecology of Organizations." *American Journal of Sociology* 82 (5): 929-64.
- Hoerr, John. 1997. *We Can't Eat Prestige: The Women Who Organized Harvard*. Philadelphia: Temple University Press.
- Hollander, Edwin P., and Lynn R. Offermann. 1990. "Power and Leadership in Organizations: Relationships in Transition." *The American Psychologist* 45 (2): 179-90.
- House, Robert J., William D. Spangler, and James Woycke. 1991. "Personality and Charisma in the U.S. Presidency: A Psychological Theory of Leader Effectiveness." *Administrative Science Quarterly* 36 (3): 364-97.
- Howell, Jane M. 1990. "Champions of Technological Innovation." *Administrative Science Quarterly* 35 (2): 317-25.
- Hutchins, Edwin. 1991. "The Social Organization of Distributed Cognition." Pp. 283-307 in *Perspective on Socially Shared Cognition*, edited by L. B. Resnick, J. M. Levine, and S. D. Teasley. Washington, D.C.: American Psychological Association.
- IUD Bulletin*. 1959. "Mitchell Hits at Conditions of American Farm Workers." Washington, D.C.: Industrial Union Department, AFL-CIO. March.
- IUD Bulletin*. 1961. "Farm Worker Organizing Urged by IUD Convention." Washington, D.C.: Industrial Union Department, AFL-CIO. December.
- Jamieson, Stuart. 1975. *Labor Unionism in American Agriculture*. Bureau of Labor Statistics, U.S. Department of Labor, 1946. Reprint, New York: Arno Press.
- Jasper, James M. 1997. *The Art of Moral Protest: Culture, Biography, and Creativity in Social Movements*. Chicago: University of Chicago Press.
- Jenkins, J. Craig. 1985. *The Politics of Insurgency: The Farm Worker Movement in the 1960s*. New York: Columbia University Press.
- Jenkins, J. Craig, and Charles Perrow. 1977. "Insurgency of the Powerless: Farm Worker Movements in the U.S." *American Sociological Review* 42:429-68.
- Kasperson, C. J. 1978. "Psychology of the Scientist: XXXVII. Scientific Creativity: A Relationship with Information Channels." *Psychological Reports* 42:691-94.

- Kim, Hyojoung, and Peter S. Bearman. 1997. "The Structure and Dynamics of Movement Participation." *American Sociological Review* 62 (February): 70-93.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry*. Princeton, N.J.: Princeton University Press.
- Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Lakoff, George, and Mark Johnson. 1980. "The Metaphorical Structure of the Human Conceptual System." *Cognitive Science* 4:195-208.
- Langer, Ellen. 1978. "Rethinking the Role of Thought in Social Interaction." In *New Directions in Attribution Research*, edited by W. Ickes, R. Kidd, and J. Harvey. Hillsdale, N.J.: Erlbaum Associates.
- . 1989. *Mindfulness*. Reading, Mass.: Addison-Wesley.
- Langer, Ellen, and L. Imber. 1979. "When Practice Makes Imperfect: Debilitating Effects of Over Learning." *Journal of Personality and Social Psychology* 37:2014-24.
- Lévi-Strauss, Claude. 1966. *The Savage Mind*, translation. Chicago: University of Chicago Press.
- Levy, Jacques. 1975. *Cesar Chavez: An Autobiography of La Causa*. New York: Norton.
- Lichtenstien, Nelson. 1968. "Walter Reuther." Pp. 280-302 in *Labor Leaders in America*, edited by Melvyn Dubovsky and Warren Van Tine. Urbana: University of Illinois Press.
- Liphart, Arend. 1971. "Comparative Politics and Comparative Method." *American Political Science Review* 65 (3): 682-98.
- Lipsky, Michael. 1968. "Protest as a Political Resource." *American Political Science Review* 62 (48): 1144-58.
- Little, Daniel. 1991. *Varieties of Social Explanation*. Boulder, Colo.: Westview Press.
- Lofland, John. 1996. *Social Movement Organizations*. New York: Walter deGruyter.
- London, Joan, and Henry Anderson. 1970. *So Shall Ye Reap*. New York: Thomas Y. Crowell.
- MacKinnon, D. W. 1965. "Personality and the Realization of Creative Potential." *American Psychologist* 20 (2): 273-81.
- Majka, Theo J., and Linda C. Majka. 1982. *Farm Workers, Agribusiness and the State*. Philadelphia: Temple University Press.
- Mannheim, Karl. 1952. "The Problem of Generations." Pp. 276-322 in *Essays on the Sociology of Knowledge*, edited by Ernest Mannheim. London: Routledge & Paul.
- Mansbridge, Jane. 1986. *Why We Lost the ERA*. Chicago: University of Chicago Press.
- March, James, and Johan Olsen. 1976. *Ambiguity and Choice in Organizations*. Bergen: Universitetsforlaget.
- Martin, Phillip, and David S. North. 1984. "Nonimmigrant Aliens in American Agriculture." Pp. 168-93 in *Seasonal Agricultural Labor Markets in the U.S.* Ames: Iowa State University Press.
- Mathiessen, Peter. 1969. *Sal Si Puedes: Cesar Chavez and the New American Revolution*. New York: Dell.
- McAdam, Doug. 1983. "Tactical Innovations and the Pace of Insurgency." *American Sociological Review* 48:735-54.
- McAdam, Doug, John McCarthy, and Meyer Zald. 1996. *Comparative Perspectives on Social Movements*. New York: Cambridge University Press.
- McAdam, Doug, and Ronnelle Paulsen. 1993. "Specifying the Relationships between Social Ties and Activism." *American Journal of Sociology* 98:640-67.
- McCarthy, John D., and Meyer N. Zald. 1977. "Resource Mobilization and Social Movements: A Partial Theory." *American Journal of Sociology* 8 (6): 1212-41.
- McCleod, P. L. 1992. "The Effects of Ethnic Diversity on Idea Generation in Small Groups." Paper presented at the Academy of Management Convention, Las Vegas.

- McWilliams, Carey. 1935. *Factories in the Field: The Story of Migratory Farm Labor in California*. Santa Barbara, Calif.: Peregrine.
- Meindl, James R. 1989. "Managing to Be Fair: An Exploration of Values, Motives, and Leadership." *Administrative Science Quarterly* 34 (2): 252-77.
- Meister, Dick, and Anne Loftis. 1977. *A Long Time Coming: The Struggle to Unionize America's Farm Workers*. New York: Macmillan.
- Metzler, William. 1964. *The Farm Worker in a Changing Agriculture*. Giannini Foundation research report no. 277. University of California, Berkeley, Division of Agriculture and Natural Resources.
- Meyer, John P., and Natalie J. Allen. 1997. "Meaning of Commitment." Pp. 8-22 in *Commitment in the Workplace: Theory, Research and Application*. Thousand Oaks, Calif.: Sage Publications.
- Miller, James J. 1961. "Legal and Economic History of the Secondary Boycott." *Labor Law Journal* (August): 751-59.
- Mills, C. Wright. 1959. *The Sociological Imagination*. New York: Oxford University Press.
- Mintzberg, Henry. 1987. "Crafting Strategy." *Harvard Business Review* 65 (4): 66-76.
- . 1994. "The Rise and Fall of Strategic Planning." *Harvard Business Review* 72 (1): 107-15.
- Mintzberg, Henry, and Alexandra McHugh. 1985. "Strategy Formation in an Adhocracy." *Administrative Science Quarterly* 30 (2): 160-98.
- Mitchell, H. L. 1979. *Mean Things Happening in This Land: The Life and Times of H. L. Mitchell Co-Founder of the Southern Tenant Farmers Union*. Montclair, N. J.: Allanheld, Osmun.
- Moore, Mark H. 1995. *Creating Public Value: Strategic Management in Government*. Cambridge, Mass.: Harvard University Press.
- Morris, Aldon D. 1984. *Origins of the Civil Rights Movement: Black Communities Organizing for Change*. New York: Free Press.
- . 1993. "Birmingham Confrontation Reconsidered: An Analysis of the Dynamics and Tactics of Mobilization." *American Sociological Review* 58:621-36.
- Moscovici, S. 1984. "The Phenomenon of Social Representations." In *Social Representation*, edited by S. Moscovici and R. M. Moscovici Farr. New York: Cambridge University Press.
- Mowday, Richard, Lyman W. Porter, and Richard M. Steers. 1982. "Nature of Organizational Commitment." Pp. 19-49 in *Employee-Organization Linkages: The Psychology of Commitment, Absenteeism and Turnover*. New York: Academic Press.
- National Advisory Committee on Farm Labor. 1967. *Farm Labor Organizing 1905-1967, A Brief History*. New York: self-published.
- Nelson, Eugene. 1966. *Huelga: The First Hundred Days of the Delano Grape Strike*. Delano, Calif.: Farm Worker Press.
- Nemeth, C. J., and B. M. Staw. 1989. "The Tradeoffs of Social Control and Innovation in Groups and Organizations." Pp. 722-30 in *Advances in Experimental Social Psychology*, vol. 22. Edited by L. Berkowitz. New York: Academic Press.
- Newell, A., J. Shaw, and H. Simon. 1962. "The Processes of Creative Thinking." In *Contemporary Approaches to Creative Thinking*, edited by G. Terrell, H. Gruber, and M. Wertheimer. New York: Atherton Press.
- Oberschall, Anthony. 1973. "Mobilization: Participation in Opposition Movements, Leaders and Activists, Opposition Ideas." Pp. 146-95 in *Social Conflict and Social Movements*. Englewood Cliffs, N.J.: Prentice-Hall.
- Oliver, Christine. 1988. "The Collective Strategy Framework: An Application to Competing Predictions of Isomorphism." *Administrative Science Quarterly* 33 (4): 543-61.

- Oliver, Pamela E., and Gerald Marwell. 1992. "Mobilizing Technologies for Collective Action." Pp. 251-72 in *Frontiers of Social Movement Theory*, edited by Aldon D. Morris and Carol McClurg Mueller. New Haven, Conn.: Yale University Press.
- Osborn, A. 1963. *Applied Imagination: Principles and Procedures of Creative Thinking*. New York: Scribners.
- Parry, K. W. 1998. "Grounded Theory and Social Process: A New Direction for Leadership Research." *Leadership Quarterly* 9 (1): 85-105.
- Perrow, Charles. 1986. *Complex Organizations: A Critical Essay*. New York: McGraw-Hill.
- Peterson, Jordan. 1999. *Maps of Meaning: The Architecture of Belief*. New York: Routledge.
- Pfeffer, Jeffrey, and Gerald Salancik. 1978. *The External Control of Organizations: A Resource Dependence Perspective*. New York: Harper & Row.
- Piore, Michael. 1995. *Beyond Individualism*. Cambridge, Mass.: Harvard University Press.
- Porter, Michael E. 1996. "Making Strategy." *Harvard Business Review* 74 (6): 61-77.
- Powell, Walter W. 1988. "Institutional Effects on Organizational Structure and Performance." Pp. 115-36 in *Institutional Patterns and Organizations*, edited by Lynne G. Zucker. Cambridge, Mass.: Ballinger.
- Powell, Walter W., and Paul J. DiMaggio. 1991. "Introduction." Pp. 1-38 in *The New Institutionalism in Organizational Analysis*. Chicago: University of Chicago Press.
- Prentky, R. A. 1980. *Creativity and Psychopathology*. New York: Praeger.
- Putnam, Robert. 1993. *Making Democracy Work*. Princeton, N.J.: Princeton University Press.
- Quinn, E. 1980. "Creativity and Cognitive Complexity." *Social Behavior and Personality* 8:213-15.
- Quinn, J. B. 1980. *Strategies for Change: Logical Incrementalism*. Homewood, Ill.: Richard D. Irwin.
- Reuther, Victor. 1976. *The Brothers Reuther*. New York: Houghton-Mifflin.
- Rickards, T., and B. L. Freedman. 1978. "Procedures for Managers in Idea-Deficient Situations: Examination of Brainstorming Approaches." *Journal of Management Studies* 15:43-55.
- Rogers, Everett. 1995. "Innovations in Organizations." Pp. 371-402 in *Diffusion of Innovations*. New York: Free Press.
- Rosaldo, Renato. 1989. *Culture and Truth: The Remaking of Social Analysis*. Boston: Beacon Press.
- Rose, Margaret. 1995. "'Woman Power Will Stop Those Grapes': Chicana Organizers and Middle-Class Female Supporters in the Farm Workers Grape Boycott in Philadelphia, 1969-70." *Journal of Women's History* (Winter): 6-32.
- Ross, Fred. 1989. *Conquering Goliath: Cesar Chavez at the Beginning*. Keene, Calif.: United Farm Workers.
- Ross, Robert J. 1983. "Generational Change and Primary Groups in a Social Movement." Pp. 177-87 in *Social Movements of the Sixties and Seventies*, edited by Joe Freeman. New York: Longman.
- Ruscio, J., D. Whitney, and T. M. Amabile. 1995. *How Do Motivation and Task Behaviors Affect Creativity? An Investigation in Three Domains*. Waltham, Mass.: Brandeis University.
- Salancik, G. R., and J. Pfeffer. 1977. "Who Gets Power—and How They Hold On to It: A Strategic Contingency Model of Power." *Organizational Dynamics* 2 (21): 2-21.
- Scharlin, Craig, and Lilia V. Villanueva. 1992. *Phillip Vera Cruz: A Personal History of Filipino Immigrants and the Farmworkers Movement*, edited by Glenn Omatsu and Augusto Espiritu. Burbank, Calif.: Freemont Press.

- Schelling, Thomas C. 1960. *The Strategy of Conflict*. Cambridge, Mass.: Harvard University Press.
- Senge, Peter. 1990. *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York: Doubleday.
- Sewell, William, Jr. 1996. "Three Temporalities: Toward an Eventful Sociology." Pp. 245–80 in *The Historic Turn in the Human Sciences*, edited by Terrence J. McDonald. Ann Arbor: University of Michigan Press.
- Simonton, D. K. 1988. "Creativity, Leadership and Chance." Pp. 386–426 in *The Nature of Creativity: Contemporary Psychological Perspectives*, edited by R. J. Sternberg. Cambridge: Cambridge University Press.
- Skocpol, Theda. 1984. "Emerging Agendas and Recurrent Strategies in Historical Sociology." Pp. 356–91 in *Vision and Method in Historical Sociology*, edited by Theda Skocpol. Cambridge: Cambridge University Press.
- Smelser, Neil J. 1962. *Theory of Collective Action*. New York: Free Press.
- Smith, Sydney D. 1987. *Grapes of Conflict*. Pasadena, Calif.: Hope Publishing House.
- Snow, David A., E. Burke Rochford, Jr., Steven K. Worden, and Robert D. Benford. 1986. "Frame Alignment Processes, Micromobilization, and Movement Participation." *American Sociological Review* 51:464–81.
- Sosnick, Stephen H. 1978. *Hired Hands: Seasonal Farm Workers in the United States*. Santa Barbara, Calif.: McNally & Loftin, West.
- Stark, Rodney, and William Bainbridge. 1985. *The Future of Religion: Secularization, Revival and Cult Formation*. Berkeley and Los Angeles: University of California Press.
- Stein, M. I. 1975. *Stimulating Creativity*, vols. 1 and 2. New York: Academic Press.
- Stinchcombe, Arthur. 1965. "Social Structure and Organizations." Pp. 143–53 in *Handbook of Organizations*, edited by James G. March. Chicago: Rand McNally.
- Strang, David, and John Meyer. 1994. "Institutional Conditions for Diffusion." Pp. 100–112 in *Institutional Environments and Organizations*, edited by W. Richard Scott and John W. Meyer. Beverly Hills, Calif.: Sage.
- Swidler, Ann. 1986. "Culture in Action: Symbols and Strategies." *American Sociological Review* 51 (2): 273–86.
- Taylor, Ronald B. 1975. *Chavez and the Farm Workers*. Boston: Beacon Press.
- Taylor, Verta, and Nancy E. Whitter. 1992. "Collective Identity in Social Movement Communities: Lesbian Feminist Mobilization." Pp. 104–30 in *Frontiers in Social Movement Theory*, edited by A. D. Morris and C. M. Mueller. New Haven, Conn.: Yale University Press.
- Tilly, Charles. 1981. "Introduction." Pp. 13–25 in *Class Conflict and Collective Action*, edited by L. A. Tilly and C. Tilly. Beverly Hill, Calif.: Sage.
- . 1986. *The Contentious French*. Cambridge, Mass.: Harvard University Press.
- Turner, Victor. 1966. *The Ritual Process: Structure and Anti-Structure*. Ithaca, N.Y.: Cornell University Press.
- U. S. Bureau of the Census. 1968. Census of Agriculture, 1964. Statistics for the State and Counties, California, vol. 1, pt. 48. Washington, D.C.: Government Printing Office.
- Villarejo, Don. 1997. E-mail communication with the author, July 21.
- Walberg, H. J. 1971. "Varieties of Adolescent Creativity and the High School Environment." *Exceptional Children* 38:111–16.
- Watson, Don. 1999. Private communication. Larry Itliong files, United Farm Workers collection, Walter P. Reuther Library, Wayne State University.
- Watson, William. 1990. "Strategy, Rationality, and Inference: The Possibility of Symbolic Performances." *Sociology* 24 (3): 480–514.
- Weber, Max. (1948) 1978. *Economy and Society*, vol. 1. Edited by Guenther Roth and Claus Wittich. Berkeley: University of California Press.

- Weick, Karl E. 1979. "An Introduction to Organizing." Pp. 1-23 in *The Social Psychology of Organizing*. New York: McGraw-Hill.
- . 1993. "Sensemaking in Organizations: Small Structures with Large Consequences." Pp. 10-37 in *Social Psychology in Organizations*, edited J. K. Murnighan. New York: Prentice Hall.
- Weil, David. 1994. *Turning the Tide*. New York: Lexington Books.
- White, H. C. 1992. *Identity and Control: A Structural Theory of Social Action*. Princeton, N.J.: Princeton University Press.
- Wickelgren, W. A. 1979. *Cognitive Psychology*. Englewood Cliffs, N.J.: Prentice-Hall.
- Wickham-Crowley, Timothy. 1992. *Guerrillas and Revolution in Latin America: A Comparative Study of Insurgents and Regimes since 1956*. Princeton, N.J.: Princeton University Press.
- Zaltman, Gerald, Robert Duncan, and Jonny Holbek. 1973. *Innovations and Organizations*. New York: John Wiley & Sons.
- Zerubavel, Eviatar. 1997. *Social Mindscapes: An Invitation to Cognitive Sociology*. Cambridge, Mass.: Harvard University Press.
- Zieger, Robert H. 1987. "George Meany: Labor's Organization Man." Pp. 324-49 in *Labor Leaders in America*, edited by Melvyn Dubovsky and Warren Van Tine. Urbana: University of Illinois Press.
- . 1995. *The CIO, 1935-1955*. Chapel Hill: University of North Carolina Press.

Interviews

- Anderson, Henry. 1996. Interview by author. Research director, Agricultural Workers Organizing Committee, AFL-CIO, 1959-62. Oakland, California, August.
- Cohen, Jerome. 1995, 1998. Interview by author. General counsel, United Farm Workers, 1967-81. Carmel, California, July and August.
- Chatfield, Leroy. 1996. Interview by author. Administrative assistant to Cesar Chavez, director of National Farm Worker Service Center, 1965-1974; administrative assistant to Gov. Edmund G. Brown, Jr., 1974-75; member, California Agricultural Relations Labor Board, 1975-77. Sacramento, California, August.
- Drake, Jim. 1997. Interview by author. Minister-organizer, California Migrant Ministry, 1962-65; boycott coordinator, United Farm Workers, 1965-66; administrative assistant to Cesar Chavez, 1966-68; New York boycott director, 1969-71; organizing director, 1971-72; New York boycott director, 1973-75; organizer, 1975-77. Boston, Massachusetts, June.
- Hartmire, Chris. 1996. Interview by author. Director, California Migrant Ministry, National Farm Worker Ministry, 1961-81. Sacramento, California, August.
- Hernandez, Julio. 1995. Transcript of an oral history interview. California State University, Northridge. Provosts Committee on Chicano Labor, March 9.
- Levitt, Herman. 1996. Interview by author. Secretary-treasurer, Hotel and Restaurant Employees Union, Los Angeles Council, 1960-96; national secretary-treasurer, 1982-96. Los Angeles, California, August.
- Medina, Eliseo. 1998. Interview by author. Organizer, 1966-77; Chicago boycott director, 1967-70, 1973-75; National Executive Board member, 1973-77; vice president, 1977-78. Los Angeles, California, August.
- Padilla, Gilbert. 1999. Telephone interview by author. Vice president, United Farm Workers, 1962-73; secretary-treasurer, United Farm Workers, 1973-80. Fresno, California, March.
- Schrade, Paul. 1984. Interview by author. Regional director, region 6, United Auto Workers, 1963-71. Los Angeles, California, July.

The Outcomes of Homeless Mobilization: The Influence of Organization, Disruption, Political Mediation, and Framing¹

Daniel M. Cress

University of Colorado at Boulder

David A. Snow

University of Arizona, Tucson

This article contributes to a more systematic understanding of movement outcomes by analyzing how organizational, tactical, political, and framing variables interact and combine to account for differences in the outcomes attained by 15 homeless social movement organizations (SMOs) active in eight U.S. cities. Using qualitative comparative analysis to assess ethnographically derived data on the 15 SMOs, the study highlights the importance of organizational viability and the rhetorical quality of diagnostic and prognostic frames for securing outcomes while identifying a contingent relationship between tactics and political environment. The analysis suggests that there are multiple pathways leading to movement outcome attainment, and therefore unidimensional rather than combinatorial and interactive approaches are misguided.

One of the major rationales for studying social movements is the belief that they have important consequences or effects typically conceptualized as outcomes.² Yet, relative to other movement processes—such as emergence, recruitment and participation, and tactical development—our un-

¹ This is an expanded and revised version of a paper that was presented at the 1997 Annual Meetings of the American Sociological Association. The research on which the article is based was supported by a grant from the National Science Foundation (SES 9008809, David A. Snow, principal investigator). We wish to thank Theron Quist and Kelly Smith for their assistance with a larger project from which the article derives and Pamela Oliver, Fred Pampel, and Yvonne Zylan for their constructive and helpful comments. We also acknowledge the assistance of the *AJS* reviewers. Direct correspondence to Daniel M. Cress, Department of Sociology, University of Colorado, Boulder, Colorado 80309.

² Throughout the article, we use the terms “outcomes” and “consequences” interchangeably, although we focus, as we indicate later, on the direct, intended consequences of movement activity.

derstanding of the consequences of social movements is conspicuously underdeveloped (Amenta, Tamarelli, and Young 1996; Giugni 1998; McAdam, McCarthy, and Zald 1988). This lacuna is not due to disinterest or inattention. As noted in a recent review essay on movement outcomes, "there exists a considerable amount of work on this" topic, but "little systematic research has been done so far" (Giugni 1998, p. 371).³ Furthermore, the development of a systematic understanding of movement outcomes is hampered by conceptual and causal confusion (Amenta and Young 1999; Burstein 1999; Diani 1997; Tarrow 1994). Conceptually, the range of outcomes attributed to social movements varies widely, extending from state-level policy decisions to expansion of a movement's social capital to changes in participants' biographies. Evaluating what counts as an outcome clearly is open to debate (Amenta and Young 1998; Diani 1997). On a causal level, the precise influence of social movement activity in relation to specifiable outcomes is difficult to ascertain (Giugni 1998). In addition, there is debate about which factors associated with social movements are most important in affecting the relative success of their outcome attainment efforts, with most approaches emphasizing the importance of one factor or set of conditions over others (Giugni 1998). And finally, the potential influence of cultural and ideational factors in the determination of movement outcomes has been glaringly absent in most theoretical discussions and research explorations of the problem.

Taking these lacunae and shortcomings into account, we seek to contribute to a more systematic understanding of social movement outcomes by drawing on our field research on the mobilization and protest activities of 15 homeless social movement organizations (SMOs) that were active in eight U.S. cities from the mid-1980s through the early 1990s. Toward that end, we first provide a conceptual framework for understanding movement outcomes based on the work of other scholars and the pursuits of the homeless SMOs. Second, we discuss and operationalize relevant organizational and political factors suggested by the three dominant perspectives on the determinants of social movement outcomes (Amenta, Caruthers, and Zylan 1992; Amenta, Dunleavy, and Bernstein 1994; Amenta et al. 1996; Gamson 1990; Piven and Cloward 1977). We then extend these perspectives by including factors associated with the framing activities of the homeless SMOs in our assessment of their outcomes (Snow, Rochford, Worden, and Benford 1986; Snow and Benford 1988, 1992). And third, using qualitative comparative analysis (Ragin 1987), we evaluate the ways

³ Giugni (1998) provides the most recent and exhaustive overview of the movement outcomes literature. Other overviews that summarize the state of thinking on movement outcomes at different moments include Jenkins (1981), Marx and Wood (1975), and McAdam et al. (1988).

in which the organizational, tactical, political, and framing factors interact and combine to generate the various outcomes pursued by the homeless SMOs. The findings indicate that although there are multiple pathways of conditions leading to outcome attainment, the viability of movement organizations and the rhetorical quality of their diagnostic and prognostic frames are particularly influential conditions, while the influence of tactics and the political environment are contingent on how they interact and combine. The conceptual and causal implications of these findings are examined for understanding social movement outcomes more broadly.

CONCEPTUALIZING SOCIAL MOVEMENT OUTCOMES

The literature concerned with the consequences of social movements generally has focused on two categories of outcomes: direct outcomes, such as securing constituent benefits and winning new advantages from targets of influence (Burstein 1998; Gamson 1990; Issac and Kelly 1981; Piven and Cloward 1977), and indirect outcomes, such as changes in public perception regarding the issue in question, the generation of countermovements, and biographical changes, including the creation of career activists (Gusfield 1981; McAdam 1988; Zald and Useem 1987). Whereas direct outcomes are typically articulated as movement goals and are a reflection of a movement's primary ideological rationale, indirect outcomes are thought to reflect a movement's influence but are less likely to be ideologically based or articulated as proximate objectives. In this article, we focus mainly on what the homeless SMOs were able to obtain from the targets of their actions and thus speak primarily to the literature on direct rather than indirect outcomes.

Direct Outcomes

The point of departure for discussion of direct outcomes is Gamson's (1990) study of the success and failure of 53 SMOs in the United States between 1800 and 1945. Gamson examined whether these challenging groups received new advantages or acceptance. New advantages encompass constituent benefits that result from movement action and demands. Acceptance results when the challenging group or SMO is viewed by its targets as a representative of a legitimate set of interests; "it involves a change from hostility or indifference to a more positive relationship" (Gamson 1990, p. 31). The cross-classification of these two general outcomes, in terms of their presence or absence, yields four more specific outcomes: full response, co-optation, preemption, or collapse.

While this framework remains the most commonly used by social movement scholars (Tarrow 1994), it is not without its critics. The work of

Piven and Cloward (1977) challenges the inclusion of organizational acceptance as evidence of success. They argue that building mass organizations is detrimental to the poor and that acceptance matters little if beneficiaries do not gain anything directly. Subsequent work on movement outcomes has prioritized new advantages over acceptance (Amenta et al. 1992) and has examined acceptance as a condition for receiving new advantages (Ragin 1989). And others have argued for the inclusion of organizational survival as an indicator of success net of goal attainment, in large part because the goals of SMOs are often elusive, achieved in increments, and redefined over time. From this vantage point, organizational survival implies progress, if not attainment of goals (Minkoff 1993; Zald and Ash 1966).

In addition to the debate over whether to include indicators of organizational success, Gamson's criteria for new advantages have been challenged. Amenta et al. (1996) argue that the achievement of SMO goals is less relevant than whether those goals actually benefit challenging groups. They emphasize collective benefits, from which nonparticipants cannot be excluded, over the achievement of a stated program. Thus, benefits received only by participants would not be considered as indicators of success even if they were among the SMO's goals. At the same time, challengers who obtained only some of their goals would not necessarily be considered failures. Relating this to the Townsend Movement, Amenta and his colleagues argue that: "The failure of the Townsend Plan does not imply the failure of the Townsend Movement" (1996, p. 3).

Representation, Resources, Rights, and Relief

Our conceptualization of movement outcomes, outlined in table 1, is based on both the work discussed above and on the range of outcomes sought by the homeless SMOs. We identify two categories of outcomes sought and obtained by homeless SMOs: organizational and beneficiary outcomes. While acknowledging the debate over the importance of organizational outcomes, we include them because they were actively sought by the homeless SMOs for good reason: when obtained, they helped the homeless SMOs by providing them with some degree of institutional voice as well as resources necessary to engage in collective action.

Thus, we identified two kinds of organizational outcomes: representation and resources. *Representation* refers to formal participation of SMO members on the boards and committees of organizations that are the targets of influence. For the homeless movement, it is indicated by homeless SMO operatives assuming positions on social service boards and city task forces that deal with the homeless issue. It is a more restrictive indicator of organizational acceptance than Gamson's (1990), which includes con-

TABLE 1
TYPES OF OUTCOMES SOUGHT BY HOMELESS SMOs

Outcome Recipient	Type	Examples
SMO	Representation	Position on city task forces addressing the homeless issue
	Resources	Position on service provider boards Office space and supplies Control of service provider organizations
Beneficiary	Rights	Securing the right to vote, go to school, and obtain welfare benefits
		Reducing or eliminating police harassment
		Reducing or eliminating merchant and service provider discrimination
	Relief	Securing accommodative facilities (shelter, soup kitchens, storage, showers)
		Securing restorative programs and facilities (employment programs, permanent housing)

sultation, negotiation, and recognition, as well as a form of representation he calls inclusion. We prefer the term representation, however, because it indicates not only formal involvement, but involvement for the purposes of representing the interests of a typically excluded and voiceless constituency. Thus, in the case of the homeless, representation provides a level of institutional voice for the homeless population by enabling the SMO to have input in policy decisions that affect homeless people.

Resources refer to material concessions received by homeless SMOs from the targets of their collective action. These material concessions sometimes included money but most often consisted of less fungible resources such as office space and supplies. As we have noted elsewhere (Cress and Snow 1996), material resources, as well as other types of resources—such as moral, informational, and human resources—were more often received from various facilitative organizations. Nonetheless, homeless SMOs often attempted to gain material concessions as part of collective action settlements with their antagonists. As with all of the resources the homeless SMOs depended on, these material ones facilitated their survival and enabled them to continue as an organizational entity to press for change on behalf of their homeless beneficiaries.

Improving the conditions of their beneficiaries—the local homeless population—was the primary objective of the homeless SMOs, with the organizational outcomes constituting a means to that end. Two types of beneficiary outcomes were routinely sought: rights and relief. *Rights* encompass outcomes that protect homeless people from discriminatory prac-

tices based on their impoverished status and those that acknowledge their citizenship. They typically included protection from police harassment and merchant discrimination as well as securing their right to vote and apply for welfare benefits. These types of practices, because they affected the everyday world of homeless people, were often the original issues around which homeless people mobilized. *Relief* refers to outcomes that help ameliorate the conditions of homelessness. More concretely, it is constituted by the provision of the basic necessities that accommodate daily survival on the streets and the creation of restorative facilities that enhance the chances of getting off the streets. Examples of accommodative relief include the provision of shelters, soup kitchens, showers, and restrooms; examples of restorative relief include jobs, job training, transitional housing, and more permanent low-income housing (Snow and Anderson 1993).

This typology of outcomes is consistent with the work mentioned above that emphasizes organizational and beneficiary success. In addition, we distinguish important subtypes of outcomes for both organizations and beneficiaries. Although these outcomes are grounded empirically in the pursuits of homeless SMOs, we think they are of broader generality and can be applied to other movement contexts. In addition, this range of outcome types allows us to think more systematically about outcomes as a dependent variable and about the factors associated with its variation across SMOs in the same family or sector. We turn now to this issue.

FACTORS ASSOCIATED WITH OUTCOME ATTAINMENT

While a number of studies have discussed the outcomes that particular social movements were able to achieve, there are few systematic attempts at theorizing social movement outcomes in general. The primary frameworks again come from the work of Gamson (1990; Gamson and Schmeidler 1984), Piven and Cloward (1977, 1992; Cloward and Piven 1984), and Amenta and his colleagues (1992, 1994, 1996), with attention focused on a movement's organizational characteristics, tactical repertoires, and political context.

Organizational Characteristics and Tactics

Gamson's (1990) analysis focuses predominantly on organizational characteristics, such as structure, goals, and tactics, to explain SMO success and failure. He argues that challengers that have single-issue and nondisplacement goals, provide selective incentives, are bureaucratic and centralized in structure, and use disruptive tactics are more likely to be successful. Subsequent reanalysis of his data generally has upheld his findings

(Frey, Dietz, and Kalof 1992; Mirowsky and Ross 1981; Ragin 1987; Steddy and Foley 1979).⁴

However, Piven and Cloward (1977, 1992; Cloward and Piven 1984) contest Gamson's emphasis on the importance of organizational characteristics for success, arguing that, at least for movements of the poor, organization-building is typically antithetical to their interests. Elites, they assert, respond not to organization, but to disruption of significant social institutions. The opportunities for effective collective action by the poor are limited to times of widespread discontent when there is a division among elites. In these instances, certain elites may ally themselves with the concerns of the poor to shore up their own power base, ultimately helping to legitimate the claims of the poor. Disruption of significant social institutions in these contexts, which are typically short lived, is what ultimately leads to concessions. Thus, Piven and Cloward contend that an emphasis on organization-building deflects energy from those moments when disruptive action might actually win concessions.

Political Context

Amenta and his colleagues (1992, 1994, 1996) provide the most recent systematic attempt to understand the determinants of movement outcomes. They refine and build on the above debate, arguing for the presence of both strong organizations and a sympathetic political context. Regarding the latter, they propose a "political mediation model," whereby successful mobilization typically requires mediation by supportive actors in political institutions. In particular, they look at the presence of sympathetic regimes and state bureaucracies that would benefit from protest outcomes in addition to the presence of strong SMOs. In the absence of sympathetic political actors, they argue that more aggressive tactics are likely to be required by SMOs in order to obtain desired outcomes.

Problems with Major Theoretical Perspectives

The foregoing approaches to understanding the precipitants of movement outcomes place different emphases on the role of organization, tactics, and political context. Gamson's analysis highlights the roles of organizational strength and disruptive tactics on movement success. Piven and Cloward argue against the efficacy of organizational strength, emphasizing, instead, disruptive tactics and divided elites. And Amenta and his colleagues high-

⁴ For dissenting interpretations of Gamson's central findings, see Goldstone (1980) and Amenta et al. (1996).

light strong organizations and either sympathetic political regimes and bureaucracies or disruptive protest in the absence of the latter.

While these approaches provide a useful starting point for discussing factors associated with the attainment of movement outcomes, we believe they oversimplify the dynamics of outcome attainment. More specifically, there are at least four issues that are sidestepped or glossed over by these perspectives. First, each of the perspectives was developed by examining historical social movements that operated in a national context. Yet, a number of scholars have noted a shift in movement activity since the 1960s to more local arenas (Hutchinson et al. 1997; McAdam 1988). Thus, whether these factors hold for contemporary movements that operate in more localized contexts remains an empirical question.⁵

Second, the factors associated with outcome attainment have typically been analyzed in a correlate fashion, while the ways in which they interact with one another has remained less developed. Strong organizations, disruptive tactics, and sympathetic political contexts may all be associated with outcome attainment, but what is left unexplored is the ways in which they combine with one another to lead to an outcome. For example, Amenta et al. (1996) suggest that disruptive tactics may be less important in a strongly sympathetic political context, but the same context may require strong organizations. Thus, the importance of the factors does not reside solely in the strength of their association with a particular outcome, but in the more complex ways they interact with each other in relation to the attainment of various movement outcomes.⁶

This leads us to our third point. The factors associated with outcome attainment may vary in their importance depending upon the type of outcome in question. Most discussions of movement outcomes focus on only one type of outcome. For example, Piven and Cloward (1977) and Amenta and his colleagues (1992, 1994, 1996) highlight the provision of social programs by the state. How generalizable their findings are to other types of

⁵ The national/local difference also has bearing on how these factors are operationalized. For example, the organizational characteristics of centralization and bureaucracy used by Gamson (1990) for SMOs with several thousand members may not be relevant for local SMOs with a much smaller number of members. Apropos this point, Edwards and Marullo's (1995) research on peace movement organizations in the United States during the 1980s revealed that not only were most small and local, but that relatively few had "a minimal formal structure" (p. 913).

⁶ For example, Gamson's (1990) bivariate analysis suggests that each factor increases the likelihood of success for the SMO. Yet, none of the 53 SMOs in Gamson's sample had all of the organizational factors he identifies. This raises the question of the ways in which the factors may combine with one another to lead to an outcome. Apropos this concern, Ragin's (1989) reanalysis of Gamson's data using qualitative comparative analysis identified different combinations of organizational factors that led to success for the SMOs.

outcomes remains an empirical question. Gamson's (1990) work is suggestive of this point in that the organizational factors he identified differed in their influence depending upon whether the outcome was acceptance or new advantages.

Finally, while the organizational, tactical, and political mediation approaches point to a number of important factors that account for variation in movement outcomes, they do not exhaust the range of explanatory variables. In particular, they fail to consider how variation in outcome attainment might be influenced by cultural or ideational factors. One set of such factors that has been overlooked concerns the manner and extent to which the identification of targets or adversaries, the attribution of blame or responsibility, and the articulation of a plan of attack or resolution affects the attainment of desired outcomes. These factors take us to a consideration of framing processes.

Framing Processes

Framing processes are linked conceptually to the recently emergent framing perspective on collective action and social movements (Babb 1996; Gamson 1992; Snow and Benford 1988, 1992; Snow et al. 1986; Tarrow 1992). Rooted theoretically in the work of Erving Goffman, this perspective views movements not merely as carriers of existing ideas and meanings, but as signifying agents actively engaged in producing and maintaining meaning for constituents, antagonists, and bystanders. The verb framing is used to conceptualize this signifying work, which is one of the activities that SMOs and their adherents do on a regular basis. In elaborating the relevance of framing processes to movement participant mobilization, Snow and Benford (1988, p. 199) argue "that variation in the success of participant mobilization, both within and across movements, depends upon the degree to which" movements attend to the core framing tasks of diagnostic framing, prognostic framing, and motivational framing. We think it is reasonable to assume that SMOs that attend to these tasks are likely to be more successful in securing their proximate goals as well. In particular, we think that the diagnostic and prognostic framing tasks play an important but unrecognized role in the attainment of desired outcomes.

Diagnostic framing is important because it problematizes and focuses attention on an issue, helps shape how the issue is perceived, and identifies who or what is culpable, thereby identifying the targets or sources of the outcomes sought; prognostic framing is important because it stipulates specific remedies or goals for the SMO to work toward and the means or tactics for achieving these objectives. If so, then attainment of the outcomes in question in this article—representation, resources, rights, and

relief—should be partly contingent on the development of coherent and well-articulated accounts of the problems and who or what is to blame (diagnostic framing), and what needs to be done in order to remedy it (prognostic framing).

Research on a number of different movements provides suggestive examples of this proposition. For example, the shift in the diagnostic framing of automobile-related deaths from auto safety to drunk driving has had profound influence on the impact of the drunk-driving movement (Gusfield 1981; McCarthy 1994). Likewise, the career of the peace movement in the United States during the early 1980s was profoundly influenced by the prognostic frame of “nuclear freeze” (Meyer 1990; Snow and Benford 1988, 1992). Similarly, the declining resonance of the nonviolence and integration prognostic frames within the black community and the emergence of competing black power and separatist frames were in part responsible for the weakening of the Civil Rights movement (McAdam 1982).

The above argument and the illustrative cases suggest that how SMOs attend to the tasks of diagnostic and prognostic framing may be just as important as organizational, tactical, and political contextual factors in accounting for variation in movement outcomes. This assumes, of course, that SMOs are likely to vary in the extent to which they sharply articulate diagnostic and prognostic frames and that these differences are consequential for obtaining movement outcomes, *ceteris paribus*.

Thus, in the following analysis, we draw on the various perspectives discussed above by assessing the ways in which organizational, tactical, political mediation, and framing factors interact and combine to account for variation in the outcomes achieved by the 15 homeless SMOs we studied.

CONTEXT, DATA, AND METHODS

The 15 SMOs we studied were local variants of a larger social movement that surfaced in numerous cities throughout the United States in the 1980s in protest to the dramatic growth of homelessness.⁷ While this movement gained national visibility with Mitch Snyder’s 60-day fast in 1983 and peaked publicly in October 1989, when an estimated 250,000 homeless and their supporters assembled at the foot of the nation’s capitol under

⁷ While there is extensive published research on the homeless issue in the United States (Burt 1992; Rosenthal 1994; Rossi 1989; Snow and Anderson 1993; Wright 1989), there is comparatively little published research on homeless protest events or on the homeless movement in general. But see the work of Barak (1991), Rosenthal (1994), Wagner (1993), and Wright (1997) for accounts of homeless insurgency in several cities across the country.

the banner of "Housing Now!," the vast majority of homeless collective actions—such as protest rallies and marches, housing takeovers, and encampments on government property—were local in organization and focus. Moreover, the scope of this activity was extensive, with the homeless engaging in collective action in over 50 cities during the 1980s,⁸ and with over 500 protests occurring in 17 of these cities, mostly in the latter half of the decade.⁹ Even though there was an effort to coordinate some of these local mobilizations by the National Union of the Homeless that originated in 1986 in Philadelphia, and even though some 15 local SMOs counted themselves as affiliates of the National Union, the movement was primarily a locality-based, city-level phenomenon.

Because of the local character of the movement, we focused our research on homeless SMOs and protest in eight cities: Boston, Denver, Detroit, Houston, Minneapolis, Oakland, Philadelphia, and Tucson. Two factors determined the selection of these cities. First, we wanted to select cities that exhibited variation in outcomes and in the range of mobilization activity that we had identified by content-analyzing newspaper accounts of homeless collective action in 17 U.S. cities that had a daily newspaper indexed throughout the 1980s.¹⁰ Second, because our funding required that the fieldwork be conducted during a three-year period and our comparative analytic strategy required more cases than usual for an ethnography, we were constrained in terms of the amount of time and energy that could be devoted to gaining access and generating a semblance of rapport in each city. Consequently, we selected cities in which we had already established contacts with SMO leaders and activists during a previous year of pilot fieldwork in Minneapolis, Philadelphia, and Tucson.

Although the above criteria drove the selection of the eight cities, and

⁸ This figure was derived from our inspection of newspaper reports assembled through the 1980s by the NewsBank Newspaper Index, which, at the time, collected selected articles from 450 newspapers in the United States.

⁹ This figure comes from content coding the population of local daily newspapers that had been indexed between 1980 and 1992.

¹⁰ We had originally hoped to conduct a random sample from among the 50 largest U.S. cities and then use the New York Times Index and Newsbank Newspaper Index to determine the incidence and intensity of homeless collective action across the sampled cities. However, prior fieldwork in Minneapolis, Philadelphia, and Tucson, including a summer working with the programs of the National Union of the Homeless in Philadelphia, made it clear that the incidence of homeless mobilization was dramatically underrepresented by these two services. It was in light of this observation that we were forced to turn to local dailies as the basis for information on homeless mobilization and collective action across U.S. cities. Our content analysis of the 17 dailies yielded a count of over five hundred homeless protest events during the 1980s across the 17 cities, ranging from a low of 6 to a high of 83, with a mean of 30.5.

thus the 15 SMOs analyzed, it is worth noting, as indicated in table 2, that these cities are quite representative of the 50 largest U.S. cities in size, region, and homeless rate. These apparent similarities notwithstanding, our primary concern is not with generalizing to the universe of homeless SMOs, but with using our case findings to refine and extend understanding of the determinants of movement outcomes. Given the similarities and differences among our cases in terms of the causal factors and the range of outcomes obtained, they are well suited for assessing the influence of factors thought to affect outcome attainment.¹¹

Our major fieldwork objective was to map the organizational fields in which the SMOs were embedded in each city and to discern patterns of interaction, resource flows, and outcomes within these fields.¹² To accomplish this, we employed a variation of snowball sampling based on an onion/snowball strategy that began with a homeless SMO in each city and then worked outward in a layered fashion, contingent on the information and referrals secured, to supporters, antagonists, and significant bystanders in the organizational field. Thus, we began in each city with a homeless SMO with which we had already established contact, observing and participating in its meetings and protest actions and interviewing both leaders and rank-and-file members. Through these activities, we identified facilitative organizations, such as churches, activist organizations, and service providers, that provided a range of assistance to the homeless SMOs. We interviewed members of these organizations to discern the type and level of their support of the homeless SMOs (Cress and Snow 1996).

We then gathered information on the targets of homeless SMO collective actions. These included police departments, mayors' offices, city councils, service providers that were viewed by the SMOs as "pimping" the homeless issue, and federal agencies such as the Department of Housing and Urban Development. Finally, we interviewed members of organizations that, while not directly involved in the homeless protest, were identified by other organizations as having particular insight into the homeless issue and the political context in which it was embedded.

¹¹ The use of case studies to refine and extend extant theoretical positions is consistent with the recent literature exploring the rationale and uses of case studies (Burawoy 1991; Feagin, Orum, and Sjoberg 1991; Ragin 1987).

¹² By organizational fields, we refer to a set of organizations that share overlapping constituencies and interests and that recognize one another's activities as being relevant to those concerns. This is an inclusive conceptualization that encompasses all organizations with which links might be established, be they facilitative or antagonistic. This conceptualization is consistent with the institutional perspective on organizations (DiMaggio and Powell 1983) and with work on multiorganizational fields in the study of social movements (Curtis and Zurcher 1973; Klandermans 1992).

TABLE 2

CITIES AND HOMELESS SMOs USED IN THE ANALYSIS

City	Population Rank among 50 Largest Cities, 1988	Region of Country	Homeless Rate Per 10,000, 1989*	Homeless SMOs	SMO Abbreviations
Boston	19	Northeast	46.7	Boston Union of the Homeless Homefront	BUH HF
Denver	25	West	29.7	Homeless Civil Rights Project Denver Union of the Homeless	HCRP DnUH
Detroit	7	Midwest	12.2	Homeless People United	HPU
Houston	4	South	8.7	Detroit Union of the Homeless Heads Up!	DtUH HU
Minneapolis	46	Midwest	32.8	Houston Union of the Homeless Alliance of the Streets	HUH AOS
Oakland	44	West	11.7	Minneapolis Union of the Homeless People United for Economic Justice	MUH PUEJ
Philadelphia	5	Northeast	32.3	Oakland Union of the Homeless Membership Caucus	OUH MC
Tucson	35	Southwest	27.2	Philadelphia Union of the Homeless Tucson Union of the Homeless	PUH TUH

SOURCE.—This table is reprinted from D. Cress and D. Snow, "Mobilization at the Margins," *American Sociological Review* 61 (1996): 1093.

NOTE.—Other cities that compose the initial sampling frame include Atlanta, Chicago, Cleveland, District of Columbia, Honolulu, Los Angeles, New Orleans, New York, St. Louis, and San Francisco.

* Derived from Burt (1992, app. A).

The onion/snowball strategy also enabled us to gather information on six SMOs no longer in existence during the course of our fieldwork from 1989 to 1992. In each case, former members were tracked down and interviewed, and other significant organizations were sought out for additional information as well.

Throughout our field research, we used the fieldwork roles that Snow and Anderson (1993) assumed in their research on the homeless in Austin, Texas: the role of the buddy/researcher when in contact with the homeless and their SMOs; and the role of the credentialed expert when dealing with other relevant organizational actors.¹³ These fieldwork roles, coupled with the onion/snowball strategy, enabled us not only to map the contours of the relevant organizational fields for each SMO, but allowed us to triangulate our data and thereby have a number of interpretive validity checks on our various sources of information, including the claims made by those involved in homeless protests.

Ultimately, data were gathered on 15 homeless SMOs that had been active between 1984 and the end of 1992 in the eight cities, with nine of the SMOs still active during the course of our fieldwork from 1989 to 1992. The 15 homeless SMOs varied in size, ranging from a half-dozen active homeless members to those with 30 or more active members. All SMOs claimed broader support among their local homeless constituents, but they differed in their abilities to secure facilitative support, to mobilize the homeless for their collective actions, and in the outcomes they attained.

OPERATIONALIZATION AND ANALYTIC TECHNIQUES

In addition to the conceptual and theoretical issues associated with movement outcomes, there are methodological concerns about establishing the causal influence of movement activity on movement outcomes (Amenta et al. 1996; Giugni 1998). Of particular concern is the problem of determining what an outcome might have been in the absence of movement activity. Our research and analysis addresses these concerns in three ways. First, as already noted, we observed and discussed homeless mobilization with activists, allies, targets, and informed neutral observers. We thus had firsthand knowledge, as well as information from key players, that was relevant to assessing the impact of homeless SMOs and their activities on the outcomes attained. Second, because we conducted fieldwork in each

¹³ Whereas the buddy/researcher assumes a sympathetic but curious stance with respect to those being studied, the credentialed expert assumes a nonpartisan stance and embraces his or her professional identity as a means of legitimating the research inquiry. See Snow, Benford, and Anderson (1986) for a more detailed discussion of these and other fieldwork roles.

of the cities in which each of the SMOs were located, we were able either to observe the temporal relationship between movement activity and outcomes or reconstruct that relationship through the triangulation of sources discussed above. And third, our comparison of the 15 homeless SMOs across eight cities not only revealed considerable variation in outcome attainment, but enabled us to acquire a sense of whether the kinds of outcomes sought were generated or provided independent of movement activity. In light of these considerations, we are able to assess the importance of the organizational, tactical, framing, and contextual conditions for each SMO in relation to the outcomes they sought and those they obtained.

Operationalizing the Conditions

One of the primary difficulties in assessing the factors theorized to affect outcome attainment is operationalizing them in a fashion that is consistent with the literature and yet relevant to local contexts. In this section, we provide operationalizations of the causal conditions identified in the previous theoretical discussion, and we indicate whether and to what extent those conditions were modified to fit the local contexts in which the homeless SMOs operated.¹⁴

Since there was not sufficient variability among the 15 SMOs in terms of the organizational dimensions of bureaucracy and centralization, we assessed whether each of the SMOs was organizationally viable (Cress and Snow 1996). By viable, we refer to SMOs that engaged in organizational maintenance and protest activities over an extended period of time.¹⁵ We

¹⁴ We coded these conditions together, drawing on both the ethnographic and newspaper data. We began by operationalizing the theoretical factors in a fashion consistent with the homeless contexts we observed. For example, Amenta et al.'s (1992) operationalization of sympathetic allies in their discussion of the Townsend Movement looked at the presence of bureaucracies that stood to benefit from implementation of the Townsend Plan and at the presence of democratic control of state houses. In our case, we examined city councils and city bureaucracies. We then looked at each condition and evaluated whether it was present or absent for each SMO. More specific determinations for conditions that are not self-evident are provided in subsequent notes.

¹⁵ Some students of SMOs might contend that viability be conceptualized exclusively in terms of temporal persistence. Our conceptualization of viability does not ignore temporal survival, but incorporates and accents SMO activity within a temporal frame. We think this conceptualization is justified for two reasons. Most important is the fact that movements, by definition, are action oriented and that what they do should thus be weighted as heavily as their temporal persistence. Additionally, the issue of what is an adequate temporal frame may vary by the scope and objectives of an SMO. National-level SMOs may require a longer period of time to establish an active agenda at that level. Local mobilization, on the other hand, may emerge quickly and last only briefly, yet still have a significant impact. Not-in-my-backyard movements often exemplify this pattern.

operationalize *SMO viability* by reference to three factors: survival, meeting regularity, and collective action campaigns. Our indicator of survival was whether an SMO existed for one year or more.¹⁶ Next, we looked at how frequently an SMO typically met, categorizing them by whether they met at least twice a month. Finally, we examined whether SMOs planned and conducted protest campaigns that included a series of interrelated protest events. If all three conditions were met, then an SMO was classified as viable, which was the case for 7 of the 15 SMOs.

We define *disruptive tactics* as those that intentionally break laws and risk the arrest of participants, such as blockades, sit-ins, housing take-overs, and unauthorized encampments. In contrast, nondisruptive tactical action includes petitions, rallies, and demonstrations that typically have been negotiated and sanctioned in advance. Eight SMOs regularly used disruptive tactics in their collective actions.¹⁷

Sympathetic allies refer to the presence of one or more city council members who were supportive of local homeless mobilization. This was demonstrated by attending homeless SMO meetings and rallies and by taking initiatives to city agencies on behalf of the SMO. Seven of the SMOs had such allies.

The presence of *city support* refers to cities that had established agencies with the specific charge of addressing the homeless problem. Boston, Minneapolis, and Philadelphia provided significant levels of shelter for the homeless paid for with city dollars. Seven SMOs operated within these cities.

We assessed the contributions of SMO framing activities by looking for evidence of articulate and coherent diagnostic and prognostic frames. Such *diagnostic frames* clearly specify what is problematic and in need

¹⁶ We use the one-year criteria because it elicited the most accurate responses from our informants in assessing the longevity of SMOs that were not in existence while we were in the field. For example, it was easier for respondents to recall whether an SMO had been in existence for one or two years than for 10 or 15 months. Our objective with this aspect of the viability concept was to find and incorporate a temporal threshold that seemed to link SMO activity and outcome attainment. Some might question whether one year is an adequate temporal threshold, but we believe that the context in which SMOs operate must be taken into consideration. One year of sustained activity by an organization of homeless people is quite an accomplishment given the highly precarious position of most homeless people and the absence of even the most basic resources assumed by other types of SMOs. See Cress and Snow (1996) for an elaborated discussion of this issue.

¹⁷ By regular, we mean that half or more of an SMO's collective actions were disruptive. We gauged this through interviews with members of the SMOs, facilitating organizations, and other relevant actors in the organizational field. Six of the SMOs that were coded nondisruptive never engaged in disruptive protest, while one, the AOS, did so in conjunction with other homeless SMOs on one occasion.

of amelioration and identify the culpable agents or institutions. Articulate and coherent *prognostic frames* specify what needs to be done in order to remedy the diagnosed problem, such as the creation of permanent housing or the building of more shelters. Evidence of these frames came from a number of sources, including their articulation among speakers at rallies and protest events that we attended, in their thematic prominence in discussions with active SMO members, and in the media coverage of SMO homeless mobilization and their protest events.

To illustrate the difference between articulate and less than articulate framing, an SMO protesting "the homeless problem" by arguing that "the government" was to blame and that what was needed was "housing" was coded as an example of nonarticulate or unfocused framing. On the other hand, those SMOs that highlighted a specific issue, such as "shelter conditions," with specific agents at fault, such as "service providers," and that called for specific solutions, such as a city investigation into shelter operations, illustrated more highly articulate and focused diagnostic and prognostic framing and were coded as such. Seven of the SMOs were found to have articulate and coherent diagnostic frames, and 11 of them had clear and focused prognostic frames.

Qualitative Comparative Analysis

To assess how these factors affect outcomes, we use the technique of qualitative comparative analysis (Ragin 1987). Based on the logic of Boolean algebra, qualitative comparative analysis (QCA) allows for identification of the multiple and conjunctural causes of some event when comparing a relatively small number of cases. It is not simply a substitute for quantitative procedures when dealing with a small number of cases, however, as the logics of analysis are different. Quantitative approaches generalize the influence of individual variables across a number of cases and have additive and linear assumptions about the influence of variables. QCA, on the other hand, is conjunctural in its logic, examining the various ways in which specified factors interact and combine with one another to yield particular outcomes. This increases the prospect of discerning diversity and identifying different pathways that lead to an outcome of interest and thus makes this mode of analysis especially applicable to situations with complex patterns of interaction among the specified conditions. In addition, QCA simplifies analysis by dropping irrelevant factors. When two combinations that lead to an outcome are identical on all but one condition, that condition becomes irrelevant in the context of the other conditions and can be eliminated, thereby reducing two combinations into one and simplifying the analysis.

To illustrate, Amenta et al.'s (1996) political mediation thesis is sugges-

tive of two possible pathways to policy outcomes: strong SMOs in the presence of a sympathetic political environment, or strong SMOs with disruptive tactics in the absence of a sympathetic political environment. In the former case, tactics would be irrelevant in the presence of the other conditions. In the latter, disruptive tactics would become necessary (along with strong organizations) in the absence of a sympathetic political environment. Thus, QCA not only increases the prospect of discerning multiple pathways to an outcome, but it allows us to identify the simplest combinations of factors that lead to a particular outcome from the many combinations that are possible.

RESULTS AND DISCUSSION

The first step in conducting such an analysis is to discern whether the dependent and independent conditions in question are present or absent for each of the cases being compared. Table 3 summarizes that step by showing the presence and absence of the six causal conditions and four outcomes for each of the 15 homeless SMOs.¹⁸

In looking at table 3, we see that there is considerable variation across the SMOs in terms of the presence or absence of the causal conditions and the number of outcomes obtained. Five or more of the causal conditions were present for six of the SMOs, with two or fewer conditions present for seven of them. Similarly, seven SMOs achieved two or more of the outcomes, while five attained only one outcome and three failed to obtain any outcomes. It thus appears that there is a significant relationship between the number of causal conditions present and the number of outcomes obtained. And that is precisely what table 4 suggests. We are more interested, however, in the relative importance of some conditions in comparison to others and in the combinations of conditions that are necessary and sufficient for outcome attainment.

¹⁸ It is reasonable to wonder about the temporal link between the theorized causal conditions and the outcomes. Were the presumed causal conditions operative prior to the outcomes in question? Our answer is a qualified yes. For both the nine SMOs that were in existence during our fieldwork and the six that were not, we were able to establish—through the fieldwork procedures discussed earlier—whether they secured each of the four types of outcomes and whether any of the six causal conditions were operative at that time. While we cannot pinpoint temporally exactly when a condition materialized prior to the attainment of an outcome, we can assert with confidence whether the condition was present at the time in which an outcome was secured. Although it is also plausible that the attainment of some outcomes might have been influenced by the attainment of other outcomes, we did not assess this possibility largely because of the limited number of conditions that can be examined through QCA (Amenta and Young 1999).

TABLE 3
 PRESENCE OR ABSENCE OF CAUSAL CONDITIONS AND OUTCOMES FOR HOMELESS SMOs

SMO	SMO Viability	CAUSAL CONDITIONS					OUTCOMES			
		Disruptive Tactics	Sympathetic Allies	City Support	Diagnostic Frame	Prognostic Frame	Representation	Resources	Rights	Relief
PUH	1	1	1	1	1	1	1	1	1	1
AOS	1	0	1	1	1	1	1	1	1	1
OUH	1	1	1	0	1	1	1	1	1	1
TUH	1	1	1	0	1	1	1	0	1	1
PUEJ	1	1	1	1	1	1	0	0	1	1
DtUH	1	1	1	0	1	1	1	0	1	0
HCRP	1	0	0	1	1	1	1	0	1	0
BUH	0	1	0	1	0	1	0	0	0	1
DnUH	0	1	0	0	0	1	0	0	0	1
HF	0	0	1	1	0	0	0	1	0	0
HUH	0	0	0	0	0	1	0	0	0	1
HU	0	0	0	0	0	1	0	0	0	1
MUH	0	1	0	1	0	0	0	0	0	0
HPU	0	0	0	0	0	0	0	0	0	0
MC	0	0	0	0	0	0	0	0	0	0

NOTE.—“1” indicates the presence of a condition or outcome; “0” indicates its absence.

TABLE 4
NUMBER OF OUTCOMES OBTAINED BY NUMBER OF
CONDITIONS PRESENT

NUMBER OF CONDITIONS	NUMBER OF OUTCOMES		Total
	0-1	2-4	
0-2	7	0	7
3-6	1	7	8
Total	8	7	15

NOTE.— $\chi^2 = 11.71$; $df = 1$.
 $P < .001$.

Thus, in what follows, we explore the combinations of factors that lead to the four outcome types and then assess the overall impact of the SMOs by looking at the combinations of conditions that led to two or more of the outcomes. We illustrate the pathways with case material from the 15 homeless SMOs. Table 5 lists the combinations of factors that yielded representation, resources, rights, and relief for each of the SMOs, as well as the pathways leading to a significant overall impact.

Representation

Six SMOs obtained positions on boards and task forces that addressed the homeless issue. Two pathways led to this outcome. Organizational viability, diagnostic frames, and prognostic frames were necessary conditions for obtaining representation. These conditions were sufficient in combination with either disruptive tactics, where allies were present, or nondisruptive tactics, in the context of responsive city bureaucracies. The first pathway was most prominent, encompassing four of the six SMOs.

The Philadelphia Union of the Homeless (PUH) is illustrative of the first pathway to representation. It was able to stack the city task force on homelessness with homeless people. The union was one of the most active and resource rich of the SMOs after applying for and receiving a \$25,000 grant to open a shelter, the first in the nation operated by homeless people. In addition, the union counted as supporters two city council members whom they could rely on to help with homeless issues. The union's use of disruptive tactics had generated a great deal of publicity for the organization, but they sought a more institutionalized avenue for having input into the homeless problem. They blamed the homeless service provider industry for monopolizing the public policy discussion on the homeless

TABLE 5

PATHWAYS TO OUTCOMES FOR HOMELESS SMOS

Outcome Pathways	SMOs
Representation:	
VARIABLE * DISRUPT * ALLIES * DIAG * PROG+	DtUH, TUH, OUH, PUH (PUEJ)
VARIABLE * disrupt * CITY * DIAG * PROG	AOS, HCRP
Resources:	
VARIABLE * DISRUPT * ALLIES * DIAG * PROG+	OUH, PUH (DtUH, PUEJ, TUH)
VARIABLE * ALLIES * CITY * DIAG * PROG+	AOS, PUH (PUEJ)
viable * disrupt * ALLIES * CITY * diag * prog	HF
Rights:	
VARIABLE * DISRUPT * ALLIES * DIAG * PROG+	DtUH, OUH, PUEJ, PUH, TUH
VARIABLE * disrupt * CITY * DIAG * PROG	AOS, HCRP
Relief:	
VARIABLE * ALLIES * CITY * DIAG * PROG+	AOS, PUEJ, PUH
VARIABLE * DISRUPT * ALLIES * DIAG * PROG	TUH, OUH, PUEJ, PUH (DtUH)
viable * DISRUPT * allies * diag * PROG+	BUH, DtUH
viable * allies * city * diag * PROG	HU, HUH
Significant Impact:	
VARIABLE * DISRUPT * ALLIES * DIAG * PROG+	DtUH, OUH, PUEJ, PUH, TUH
VARIABLE * disrupt * CITY * DIAG * PROG	HCRP, AOS

NOTE.—Uppercase letters indicate presence of condition and lowercase letters indicate the absence of a condition. Conditions not in the equation are considered irrelevant. Multiplication signs are read as “and” while addition signs represent “or.” SMOs in parentheses are contradictory, i.e., the conditions were present but they failed to obtain the outcome.

issue. In addition, they insisted that homeless people be part of these discussions since their lives were most directly affected. The union's articulate diagnosis and prognosis of the issue is typified by one of the leaders of the Philadelphia Union of the Homeless, as he discussed their efforts to obtain representation on the Mayor's Task Force on Homelessness:

Homelessness is a fast-growing industry. Poverty pimps have sprung up all over making money off our misery, and we haven't had anything to say about it. We have to be part of the decision making that governs our lives. The Mayor's Task Force on Homelessness originally had none of us on it. Then they said Chris [the union president] could come on. Instead, we went in with 45 homeless people and got 15 of them elected on the board. Service providers lack the urgency that we have. They go home at 5 P.M. Our troubles just begin at that time.

The PUH and three other SMOs combined organizational viability and articulate framing with disruptive tactics and their council allies to obtain representation for homeless people. But note that the same conditions were present for People United for Economic Justice (PUEJ), yet it failed to secure representation.¹⁹ What accounts for its failure when four other SMOs were successful under the same conditions? We attribute its failure to a radical-flank effect (Haines 1984), as a more moderate competitor, the Alliance of the Streets (AOS), was able to secure representation for the local homeless population from the city of Minneapolis. PUEJ used disruptive tactics in a city that had been reasonably responsive to the homeless problem. This alienated PUEJ from decision makers that might otherwise have considered its input. The AOS was more likely to work with the system and use moderate tactics in its collective action. This relationship between disruption and responsive cities is underscored in the next pathway as well.

The second combination leading to representation encompassed two of the homeless SMOs. In these cases, viable SMOs with articulate diagnostic and prognostic frames used nondisruptive tactics in cities with agencies

¹⁹ In computing QCA, one needs to decide what to do with contradictory combinations—those that result in both success and failure—with regard to the outcome in question. By omitting contradictory combinations from the analysis, the results show combinations where the outcome was certain, while including contradictory combinations shows the combinations where the outcome was possible (Amenta et al. 1992). We opt for the latter strategy for two reasons. Theoretically, the former approach is overly deterministic. None of the work discussed has argued that the posited factors would ensure outcomes, only that they increased the likelihood of success. Empirically, by including contradictory combinations in our analysis, we encompass all homeless SMOs that received the outcome and thus have more cases to draw on for understanding the factors that lead to mobilization outcomes.

established to deal with the homeless problem. This pathway differed from the first one in that the SMOs used more legitimate and institutional forms of collective action in cities that had already demonstrated some attentiveness to the homeless issue. Disruptive action was considered inappropriate and potentially damaging in these contexts. We look at the Homeless Civil Rights Project (HCRP) in Boston to illustrate this pathway.

The HCRP, as its name implies, focused on issues of civil rights for the homeless. HCRP was able to obtain representation on a citizen advisory board for the local police department. In addition, it conducted workshops for police officers dealing with homeless people. This outcome, which provided a significant opportunity to influence the policing of homeless people, grew out of a campaign to free the Boston Common of a notorious police officer. The leader of HCRP discussed the situation prior to the campaign:

If you were up in the Boston Common, and if you were perceived as homeless, you were going to get kicked off the bench. . . . They had a guy. He was infamous among homeless circles; he was called Robo Cop. He was a Boston motorcycle cop, and his beat was the Commons. This guy was unbelievable. . . . He was convinced that he had the right to kick you out of the park because it was his park, and it wasn't yours. . . . When he first come onto you he'd say, "Hi guys, how you doing today? You know, geez, it's a nice day. Listen, I'll be back in 10 minutes, and you better be gone, okay?" Very nice. But when he came back 10 minutes later, if you weren't gone, he'd fuckin' manhandle you, arrest ya. And if you said, "Gee Paul, I ain't doing nothing," he'd go to a trash barrel, and he'd come out with an empty bottle, and he'd say, "public drinking." And when we first went into business, we put out flyers, and we went up to the Commons and talked to people, and we said, "you know, if we get together, we can get rid of Robo Cop." And they thought we were talking about getting rid of God.

So the HCRP put together a petition to get Robo Cop removed. This process ultimately led to negotiations with the city to gain representation on the citizen advisory committee and provide training to police officers in dealing with homeless people. The HCRP had strong resource support from a benefactor organization (Cress and Snow 1996). In addition, their framing activities were focused on civil rights issues. They identified police harassment as a particularly onerous civil rights violation and saw the ultimate solution in representation on oversight committees of the police. Finally, they were nondisruptive in their approach, using the institutional process through gathering petitions and affidavits to make their case. As we mentioned above, nondisruption appears to be more effective in cities that are sympathetic and responsive to homeless issues and SMO campaigns.

Resources

As indicated in table 5, four homeless SMOs were able to obtain material concessions for organizational use from the targets of their collective actions. At the same time, the pathways leading to resources also contained three contradictory cases, the most of any of the outcomes.²⁰ Resources were the most difficult of the four outcome types to obtain because targets were understandably reluctant to provide material support to homeless SMOs that were challenging them or making demands that might alter their current resource-allocation calculus.

Three combinations led to resource concessions, with allies on city councils a necessary condition in each. In the first pathway, allies combined with viable SMOs that had articulate diagnostic and prognostic frames and that were disruptive. The case of the Oakland Union of the Homeless (OUH) is illustrative. The union, with strong organizational support from American Friends Service Committee (AFSC), had been involved in an extensive and long-term campaign not only to get more services for the homeless, but to control and operate those services as well. The union eventually won control of a multimillion-dollar housing and service project for the homeless.²¹ Their leader discussed a segment of the collective-action campaign that resulted in this "stunning" victory:

We went over that first night, and we took over three houses and had them barricaded real tight. Then the cops came down and busted us. When our supporters found out we were locked up and the cops had arrested us for taking over the houses, they demanded that some of the council members do something on our behalf. Hours after that, we were released. We don't know how many of them pushed for it, but two city councilmen helped us out. We did that like two more times in the next two months, and the city council got totally freaked. So we kept doing these takeovers. That is the key reason why the city said, "We are willing to negotiate, what do you want?" And we said, "(permanent) housing." So out of our protest, we have a 26-unit 4.7 million-dollar construction program that breaks ground next month and will be completed in 12 months.

Thus, as a result of their organizational strength, disruptive tactics, and city council allies, the Oakland Union of the Homeless came to control extensive resources consisting of a multimillion-dollar housing project.

Yet, three other SMOs combined the same factors and failed to obtain

²⁰ It also should be noted that for some of the outcomes a number of the SMOs are implicated in more than one pathway. This is because qualitative comparative analysis reduces any set of conditions to its simplest combination. In the case of the PUH, e.g., all six conditions were present, but not all six were necessary to attain resources. Thus, it was encompassed by both pathways.

²¹ It is the control of the program and the assets that accompanied this control that qualifies the placement of this outcome into the resource category.

resources. What explains these contradictory cases? Like the OUH, both the Detroit Union of the Homeless (DtUH) and the Tucson Union of the Homeless (TUH) operated in cities that had been nonresponsive to the homeless issue. But the OUH utilized disruptive tactics more often than either the Detroit or Tucson unions. We think it was this qualitative difference in the amount of disruption, which is not fully captured by our dichotomization of tactical disruption, that accounted for the difference in resource attainment.²² As for PUEJ, it competed with a more viable and less disruptive SMO, the AOS, in Minneapolis, a city that had been responsive to the homeless problem, thus suggesting the operation of a radical-flank effect.

The second set of factors leading to resources combined viable SMOs with articulate diagnostic and prognostic framing in responsive cities where the SMO had allies on city councils. The PUH exemplified this pathway. Like the OUH, the PUH was able to gain control of homeless services, including a housing program that encompassed more than 200 housing units. One of the founding members discussed how this came about: "We were told by some people that we ought to try and get a shelter together that would be run by homeless people themselves. We applied for a grant through the city and received \$21,000. We figured we were given the money in order to fail, that the city knew we would fail and could say, 'see, we gave 'em a chance, and they fucked it up.' We had been meeting in a Methodist church as a base for our organizing, and we used this space to begin sheltering people."

The PUH had a well-articulated diagnosis of and prognosis for the homeless issue, which were embodied in its slogan, "Homeless Not Helpless." Just as with representation, the PUH recognized the homeless people had little control over the institutions that shaped their lives. Obtaining resources to run the services that homeless people depend on was seen as a significant step toward empowerment. In addition, the PUH was the oldest and most active of the homeless SMOs. As such, it had developed council allies who pushed for its initiatives.

The third pathway to resources also highlighted the importance of contextual factors, combining nonviability and nondisruptive tactics in responsive cities where the SMO had allies on city council. The city of Boston had been more active than most cities in providing shelter to the

²² This shortcoming is inherent in the QCA method, which requires the dichotomization of conditions so that they are either present or absent. At the same time, this underscores the benefits of combining qualitative comparative analysis with traditional ethnographic approaches. We are thus able to interpret contradictory cases by bringing to bear additional field data on the specific circumstances of the homeless SMOs.

homeless and its mayor had chaired the League of Cities Homeless Commission. Homefront (HF) was a Boston SMO that emerged during the course of a demonstration organized by local service providers. After the demonstration ended and the service providers and supporters went home, the homeless, who had nowhere to go, continued the demonstration and formed their organization. The support of two city council members was critical to the success of the demonstration, however. An aid to one of the city council participants explained their involvement:

It's public knowledge our office was very helpful in giving the moral and physical support for the demonstration to continue. Our staff people went down and fixed coffee for them in the morning and helped them find [photocopying] for their flyers. I myself spent weekends down here talking to them. We helped them to organize into a little government thing where they did cleaning and stuff like that. We lent logistical support. The demonstration lasted for a long time. Finally, they came to an agreement with the city, and the city gave them an office and some stuff.

Thus, the combination of active allies on the city council, a city generally responsive to the homeless issue, and nondisruptive protest eventually enabled HF to obtain an office and supplies for its operation.

Rights

Seven of the SMOs were able to secure basic rights for the homeless, such as the right to vote, to go to school, or to obtain welfare benefits, as well as protection from discriminatory practices by police, service providers, and merchants. The same combinations leading to representation also lead to rights. Again, SMO viability and diagnostic and prognostic frames were necessary conditions that, in combination with disruptive tactics and allies or the absence of disruptive tactics in responsive cities, led to the realization of one or more rights. The experiences of the DtUH illustrate the first combination.

One of the DtUH's primary concerns was to address problems of homeless children. The union was particularly troubled by the fact that in many shelters children were unable to go to school for lack of transportation. The union argued that education was a right and that the Detroit Board of Education was responsible for making sure that homeless children could get to school. One of the leaders of the union discussed its efforts to help alleviate this problem:

We have many concerns about the conditions of the shelter and the people that were there. And one of the issues that first came up was the children that were in a transient situation. Once they got to [a local shelter], there was no schooling because of [the lack of] transportation. The union, along with [a local welfare rights group], demanded and got meetings with the

Detroit Board of Education. This went on for a month's time extensively. Out of those meetings, transportation not only from [the local shelter] but to all the other shelters was arranged.

Thus, the DtUH was able to get a meeting with the school board through the help of their city council allies and the threat of disruptive tactics, which they had used regularly in other campaigns. Equally important, their diagnostic and prognostic framing activities succeeded on two levels: by targeting children the DtUH focused attention on a population that could not be blamed for its homelessness and that was most likely to be viewed in a sympathetic light; and by targeting education, it focused on an issue that the city was mandated to address. It was through the combination of these factors, then, that the DtUH ultimately succeeded in ensuring that homeless children in Detroit could continue to attend their schools.

The second combination of factors leading to rights included SMO viability and diagnostic and prognostic frames in conjunction with the absence of disruptive tactics in more responsive cities. We look at the HCRP to illustrate this pathway. One issue it addressed, in addition to police harassment discussed earlier, was merchant discrimination against homeless people. After documenting numerous cases of homeless people being refused service or told to leave restaurants while other people were allowed to linger over their meals, HCRP decided to target a national chain of coffeehouses that had a particularly notorious record of such discrimination. Its leader discussed what occurred:

We were going to set up a picket, but then we got calls from two people who were the co-chairs of the Board of Directors. They have like 200 restaurants. And they said, "Hey, we're not bad guys, can we sit down and talk?" We had been set on picketing, but we had also had an idea all along about writing up guidelines that would tell merchants how homeless people expected to be treated. Then we got to thinking, wouldn't it be better, it wouldn't be as much fun, but wouldn't it be better if they co-drafted those things with us? So that's what we asked them to do. We had a work meeting, and we hammered out the guidelines. Then we called a press conference to sign the agreement between [the chain] and the Homeless Civil Rights Project.

The subsequent publicity around the event helped to reduce the level of discrimination by local merchants. In part, this was because the approach of the HCRP showed a different side of homeless people to local merchants and also because the City of Boston retreated from its use of the police force to remove the homeless from local businesses. Thus, the presence of a strong SMO with articulate diagnostic and prognostic framing, which used nondisruptive tactics in a responsive city, led to increased protection from harassment for the local homeless population.

Relief

Relief in the form of accommodative or restorative facilities was the most widely obtained outcome. Nine SMOs were able to obtain relief within their respective cities for the local homeless population. These included such things as shelters, soup kitchens, public restrooms, showers, job programs, and permanent housing. There were multiple pathways to relief, with four combinations leading to the outcome. In addition, these combinations diverged from each other more than for the other outcomes in that only one condition was necessary to obtain relief.

The AOS in Minneapolis illustrates the first pathway to relief, which combines viability, allies, responsive cities, and articulate diagnostic and prognostic framing. The AOS led a campaign to save 150 units of low-cost housing in a downtown area of the city slated for demolition to make way for a new convention center. The AOS was the most prominent of three homeless SMOs in Minneapolis and received strong resource support from a local church, which included a clubhouse for members. The AOS's diagnostic frame blamed the city for backing out of an agreement guaranteeing one-for-one replacement for low-cost housing lost in construction, and its prognostic frame insisted that the city honor its commitment to housing the poor by building new, affordable housing. The city, in part because of its severe winters, had played an important role in sheltering homeless people. As such, the AOS had worked with and was on reasonably good terms with decision makers in the city agency dealing with the homeless issue. In addition, the AOS counted two councilmen as allies who had attended their rallies and pledged support. The AOS's leader discussed the outcome of the campaign: "We were the front lines on the Laurel Village Development project, which came out of the convention center demonstrations and a Labor Day rally that drew over fifteen hundred homeless people. Eventually, six million dollars was slated to revamp Paige Hall for low-income people in addition to the convention center housing we saved."

The second pathway to relief encompassed SMOs that combined viability and articulate framing with disruptive tactics and allies on city councils. The TUH illustrates this pathway. With significant resource support from a local Catholic Worker community, the TUH's diagnostic frame was strongly influenced by the social gospel stance of the Catholic Workers. The union argued that the presence of homeless people in the midst of affluence was something that public officials at all levels of government needed to address. Its prognosis thus included providing jobs and housing to homeless people and, short of that, improving the quality of life on the streets. The TUH utilized disruptive tactics in their protest and had allies on the local city council as well.

One particular campaign during the Christmas season typified the TUH's approach to protesting the homeless problem. The TUH organized a Posada, a traditional Mexican procession that reenacts Joseph and Mary's search for a place to stay. In the union's version, hundreds of homeless people paraded from the federal building to the city offices. The Posada, led by banners proclaiming "Still No Room at the Inn," visited various agency offices that the TUH felt should be addressing the homeless issue. The procession ended at the county building where a two-week encampment at the front of the building ensued. The county ultimately made \$50,000 available to homeless service providers to expand homeless services.

Although this combination resulted in relief for four SMOs, it was not a guarantee of relief, as the same conditions failed to generate this outcome for the DtUH. In this case, we believe that a closed political opportunity structure was particularly significant. With a strong mayor/weak council form of government, and a mayor hostile to expanding programs for the poor, the DtUH tactics and allies were less effective in securing relief than was the case for many of the other successful homeless SMOs.

The third pathway to relief encompassed SMOs that combined disruptive protest and prognostic frames with the absence of viability, allies, and diagnostic frames. The Denver Union of the Homeless (DnUH) exemplifies this pathway. This SMO had a punctuated history, with episodic bursts of protest followed by periods of dormancy. This instability prohibited it from developing and establishing connections with political allies. A dynamic and assertive leader who believed in using dramatic and disruptive protest to call attention to the homeless issue ran the DnUH. The DnUH lacked a coherent diagnostic frame that guided their actions. Rather, it used disruptive protest to call for specific actions by the city to alleviate some aspect of the homeless problem of concern to its organizer. A leader of a local activist organization that worked with the DnUH elaborated:

Dorothy (DnUH's leader) had this idea. She was setting up this shelter, she talked to some people, and she got a house down here. So we helped her get a board together, and helped to work out an agreement with the people she got the house from, and helped her start to raise some money. Well, in the process of doing that, Dorothy is looking around and saying, "Jesus, there are all these HUD [Department of Housing and Urban Development] houses with nobody living in these places. What the hell, we should be living in those. I mean, homeless people." So we talked with her and talked with her. Dorothy is the kind of person who moves on instinct, not a lot of strategy necessarily, particularly in a group sense. But individually, she decides that there are some houses not far from where her shelter is that would be excellent places for people to live. So they [the DnUH] go down and take a crowbar and rip the doors open and start spending time in the houses.

As a result of this takeover, HUD entered into negotiations with a number of local service providers and 55 houses were made available for housing for the poor. One of the HUD officials involved in the negotiations recalled the resolution to the housing takeovers:

The regional administrator at that time had made a decision that he would try to facilitate the leasing process and make more homes available more quickly in response to a need. For that reason, he called in a couple of the local providers who were already working in the program and another local church community who had called him and expressed interest, and he invited Dorothy as well. He asked the attendees if he were to make available rather quickly a number of homes, how many did they feel that they could manage, that meant financially and property management responsibility? And then each of those groups responded with a number, and it was at that time he said we will make 55 homes available to these three organizations.

This pathway suggests that sometimes very little is necessary beyond disruption and a target to get relief. In the next pathway, even disruption is unnecessary.

The final pathway to relief encompassed SMOs that had developed prognostic framing combined with the absence of viability, allies, a responsive city, and diagnostic framing. Two SMOs in Houston—Heads Up (HU) and the Houston Union of the Homeless (HUH)—were associated with this pathway. Neither SMO survived beyond a year, mainly because of the lack of resource support from other organizations. Because of this, they met irregularly and conducted only one significant event, a march and rally to bring attention to the homeless issue. Houston had done very little to address the homeless problem in the city, relying primarily on private initiative to address the issue. While the march drew approximately six hundred homeless people and their supporters, it lacked a coherent and robust diagnosis of the homeless problem. This was clearly indicated in a conversation with one of the organizers from HU: “Well, the march was one of our activities, and I wanted to support it because we needed publicity and we needed to get involved in activities which would put the Heads Up name out there. And the march itself I didn’t think was too negative. I mean, the march was just to highlight the fact that homeless people aren’t getting all the services that could be provided and to just keep the homeless issue up front.”

At the same time, both SMOs had more specific ideas of what needed to be done. An organizer with the HU discussed what it was seeking: “The march, the intent of the march was to stress the need for more low-income housing, also for emergency shelters and also to demand that the city does something about returning funds to the federal government for non-use because they don’t have programs in place to utilize the funds.”

The city responded to the march by fixing drinking fountains and set-

ting up portable toilets in a park where homeless people congregate. Thus, while this pathway suggests that relatively little is required to secure some forms of relief, the level of response also appears to be rather minimal.

Significant Impact

The above scenario underscores one of the weaknesses of our analysis: it fails to differentiate amounts of an outcome received. This is due not only to the highly variable accounting practices of the SMOs, but also because we found it more reliable to discern whether an outcome had ever been obtained by an SMO than to try and track and measure the level of the outcome received.²³ Yet, the problem of distinguishing SMOs that made a significant impact on the local homeless issue from those whose influence was more modest remains. We assess significant impact by looking at whether the SMOs received two or more of the outcome types excluding resources. We focus on representation, rights, and relief because these outcomes had the most direct impact on the local homeless population. Using this criterion, seven of the homeless SMOs were able to have a significant impact on the homeless problem in their respective communities.

As shown at the bottom of table 5, two pathways led to significant impact, with no contradictory cases in either. Once again, viable organizations with articulate diagnostic and prognostic frames were necessary conditions. These conditions combined with disruptive tactics and city council allies, or the absence of disruption in responsive cities. The first pathway encompasses five SMOs. The most successful of these were the Oakland and the Philadelphia unions. Both came to control housing assets in the multimillion-dollar range, and both expanded the accommodative resources for homeless people in their cities. In addition, they were able to protect and establish rights for homeless people around the issues of voting and welfare benefits. Finally, they were able to secure for the homeless a significant voice in policy discussions regarding the homeless problem. Thus, they were considered major players regarding poverty issues in their respective cities.

The other three SMOs in the first pathway were less effective but were still able to obtain two of the three types of outcomes and keep attention on the homeless issue. While the TUH and PUEJ never controlled housing assets, they did help save significant numbers of low-cost housing in their

²³ It also is important to keep in mind that the significance or marginality of a particular level or amount of an outcome is somewhat relative because it is highly context dependent. For example, is it better for the homeless if 150 units of low-cost housing are saved or if a new shelter is built? It might depend on whether the outcome occurred in Minneapolis or Tucson.

cities. In addition, they were both able to help change the treatment of homeless people by the police. The DtUH, on the other hand, never obtained relief for homeless people, but it did provide voice in policy debates and on task forces and fought for welfare benefits for people without an address. These accomplishments made it one of the more visible and important poor people's organizations in Detroit.

The second pathway to significant impact combined viable SMOs with articulate diagnostic and prognostic framing and the use of nondisruptive tactics in sympathetic city contexts. Both the HCRP and the AOS tended to work "with the system" whenever possible; when they protested, they tended to utilize institutional channels of redress. The AOS was able to save hundreds of units of low-cost housing and improve the conditions in local shelters. In addition, it provided homeless people with input into the decision over who should have access to permanent housing. It also helped the homeless to secure the right to vote and to obtain welfare benefits. The HCRP focused more on rights protection. As such, it was able to reduce significantly the harassment of homeless people by merchants, police, and service providers. It was also able to improve conditions inside shelters and provide homeless people with a voice on policing and shelter policies in Boston.

DISCUSSION AND IMPLICATIONS

We noted at the outset that, relative to other social movement processes, there is a paucity of research that attempts to systematically understand the outcomes of social movements. These efforts have been hampered by both conceptual and causal issues. We have attempted to shed additional light on these concerns by examining ethnographically the outcomes of 15 homeless SMOs in eight U.S. cities. Specifically, we identified the range of outcomes pursued and obtained by the homeless SMOs, discussed and operationalized four sets of explanatory conditions, and identified the conjunction of factors that led to attainment of various combinations of the outcomes pursued across the 15 SMOs. We summarize our findings below in the course of discussing their relevance for thinking about social movement outcomes more broadly and their implications for understanding more thoroughly the causal dynamics underlying the attainment of movement outcomes.

Conceptualizing Social Movement Outcomes

One implication stems from our identification of the range of outcome types pursued by the homeless SMOs. Much of the discussion of movement outcomes emphasizes the passage of governmental policy or legisla-

tion (Giugni 1998), ignoring or treating superficially the various outcomes we have identified. The homeless SMOs we studied rarely attempted to pass legislation, however. This is not to suggest that there was disinterest in national legislation that benefited the homeless. In fact, homeless SMOs and activists from around the country have coalesced on a number of occasions to appeal to the federal government for assistance in dealing with the problem of homelessness. A case in point was the previously mentioned 1989 "Housing Now!" march on Washington D.C. In general, however, the homeless SMOs we studied, as well as others we have learned about secondhand, worked to achieve more proximate goals, such as getting homeless people off the streets, and short of that, to improve the quality of life on the streets for those unfortunate enough to become homeless. In addition, they sought to represent the voice of homeless people in policy discussions and enrich their organizations in order to further their mobilizing capacity. The kinds of outcomes we identified were significant not only to those homeless individuals whose lives benefited concretely from them, but to the ongoing mobilization of homeless protest in general.

We think the emphasis on social policy outcomes is largely a function of the national-level emphasis of much social movement scholarship rather than the empirically demonstrated result of what most social movements actually attempt to accomplish. Researchers have tended to study large and presumably successful movements at the expense of more localized or failed ones (Voss 1996), and the passage of national policy is often seen as the pinnacle of movement impact. Yet, even movements that appear to have a national focus tend to have a broader range of concerns that we believe are captured by our typology. For example, research on the Civil Rights movement clearly shows how SMOs worked to extend rights, secure various forms of relief, and represent the voices of blacks in policy discussions (McAdam 1982; Morris 1984), as well as secure resources from the state for their own organizations (Haines 1984; Jenkins and Eckert 1986).

We believe the outcome types pursued by the homeless SMOs are common to other movements as well. For example, the mobilization of the Communist Party during the 1930s was built not on national policy initiatives, but through locally organized demonstrations to improve the immediate circumstances of those impoverished during the depression (Goldberg 1991). Securing welfare benefits and demanding rent relief are only a few of the several kinds of outcomes that were significant not only to those whose lives were eased, but to building the communist movement. Likewise, the main thrust of many neighborhood movements has been to provide official representation of neighborhood communities within institutional political structures.

These observations suggest that focusing on broad policy outcomes may capture only a fragment of what some, and perhaps most, SMOs actually do. It glosses over the more proximate impact that social movements can have for their beneficiaries by missing much of what is pursued in SMO collective action campaigns at a local level. Our findings and analysis provide a partial corrective to this tendency. Whether other movements operating in other contexts pursue the outcomes we identified is, of course, an empirical question. But at the very least, our outcome typology underscores the more local agendas of some varieties of SMOs and thus suggests that social movements may sometimes have greater impact than often presumed or theorized.

Causal Implications for Understanding the Dynamics of Social Movement Outcomes

One of the more general implications of our findings is that most social movement outcomes are probably obtained through multiple pathways rather than through one surefire pathway or set of conditions. As shown in table 6, which lists the multiple pathways to the outcomes of interest and their general level of impact, the four outcomes we identified were obtained through six different combinations of the six causal conditions. This suggests that the search for a single general framework that explains outcome attainment is misguided. But does it necessarily follow that each outcome requires its own unique combination of underlying conditions, thereby calling into question the utility of theoretical generalization regarding the relationship between causal conditions and outcomes? Or are there particular combinations of conditions that are more potent or efficacious than others in generating outcomes?

Looking at table 6, it would appear at first glance—with six different combinations of the six causal conditions—that the relationship between the conditions and outcomes is rather indeterminate and bereft of much theoretical utility. But closer scrutiny of the relationship between the six pathways and their respective impact indicates that such a conclusion is unwarranted. In particular, three observations stand out. First, all of the six pathways were not equally potent or successful, as measured by the number of outcomes they obtained: one pathway facilitated attainment of all four of the outcomes; two combinations led to two of the outcomes; and three of the combinations were associated with only one outcome. Thus, some pathways or combinations clearly had a more pronounced impact than others. Second, the three combinations that led to two or more outcomes were characterized by the presence of viable organizations with articulate diagnostic and prognostic framing. Where they differed was in terms of the interaction of political context and tactics. Disruptive

TABLE 6
MULTIPLE PATHWAYS TO OUTCOMES AND LEVEL OF IMPACT

Pathways	Outcomes	Impact
1. VIABLE * DISRUPT * ALLIES * DIAG * PROG	Representation, Resources, Rights, and Relief	Very strong
2. VIABLE * disrupt * CITY * DIAG * PROG	Representation and Rights	Strong
3. VIABLE * ALLIES * CITY * DIAG * PROG	Resources and Relief	Moderate
4. viable * DISRUPT * allies * diag * PROG	Relief	Weak
5. viable * allies * city * diag * PROG	Relief	Weak
6. viable * disrupt * ALLIES * CITY * diag * prog	Resources	Weak

NOTE.—Uppercase letters indicate presence of condition and lowercase letters indicate the absence of a condition. Conditions not in the equation are considered irrelevant. Multiplication signs are read as “and.”

tactics were effective in conjunction with city council allies, whereas non-disruptive tactics were more effective in responsive cities. Where both responsive cities and council allies were present, tactics were irrelevant. Third, in contrast to these basic combinations of conditions that characterized the three most consequential pathways, the remaining pathways shared few common conditions and were associated with SMOs that were idiosyncratic in terms of their origins or careers.

Based on these observations, it seems reasonable to conclude—at least for the cases we examined—that while there is some variation in the relationship between combinations of conditions and types of outcome, there are a number of combinations that clearly are more potent or efficacious in terms of the outcomes secured or attained. We examine these combinations and discuss their relevance to the outcome frameworks proposed by Gamson, Piven and Cloward, and Amenta and his colleagues, as well as to the framing perspective.

Political mediation and tactics.—The three pathways that yielded the most outcomes, and thus were most potent, extend and refine the political mediation arguments of Amenta and his colleagues (1992, 1994, 1996) and the emphasis placed on disruptive tactics by Gamson (1990) and Piven and Cloward (1977), among others. At least one of the political mediation conditions was typically required for some level of outcome attainment. In addition, the political mediation condition interacted with the use of disruptive tactics in various ways: Disruptive tactics worked in conjunction with allies on city councils, but nondisruptive tactics were necessary in more responsive cities. We think the reasons for these interactions are twofold: the presence of elite allies provides a legitimate voice for putting a positive spin on disruptive protest and for placing the SMO's grievances and objectives on a city's calendar as something that warrants positive attention and action. However, when such attention and action are present in the context of an already responsive city, disruptive protest is unnecessary and likely to be counterproductive. In turn, when both allies and a responsive context are present, protest tactics should be irrelevant, which was the case in the third pathway in table 6.

These findings call into question the generality of the significance Piven and Cloward (1977), and to a lesser extent Gamson (1990), place on disruptive tactics in relation to movement success and are more in line with the research of Amenta and his colleagues (1992, 1994, 1996), who suggest that disruption may be required only in the absence of political mediation. But even though Amenta and his colleagues provide a more robust understanding of the relationship between political mediation and disruptive tactics, they are relatively silent on the role of disruption in the presence of these conditions. Our findings refine this relationship further. In the case of the homeless SMOs, disruption was ineffective in the absence of

allies, and it was detrimental in the presence of responsive contexts. Thus, for the homeless SMOs, disruption in its own right was never sufficient to secure desired outcomes. Rather, it typically required the additional presence of allies, and, if pursued in responsive cities, could even backfire, particularly in the presence of a more moderate SMO.

Organizational viability and framing.—Our analysis also suggests that the attainment of movement goals is strongly facilitated by viable organizations that are skilled at diagnostic and prognostic framing. We argued in an earlier article that organizational viability was predicated on successful resource mobilization and suggested that viable organizations were also more likely to be successful (Cress and Snow 1996). This was consistent with Gamson's (1990) emphasis on strong, "combat ready" organizations in contrast to Piven and Cloward's (1977) argument that investment in organization-building undermines successful mobilization by diverting energy from disruptive protest. Our earlier argument was confirmed by the finding that significant outcome attainment was predicated, in part, on having viable organizations. While nonviable SMOs were sometimes able to obtain an outcome, this was much more the exception than the rule. In addition, the level of response tended to be less. For example, HU in Houston was able to obtain relief in terms of public toilets and drinking fountains. But this paled in comparison to the relief acquired by the viable SMOs, all of which created or salvaged low-cost housing worth hundreds of thousands of dollars or more.

We also have seen that the framing activities of SMOs are important for successful outcome attainment. However, framing activity associated with the development of reasonably articulate and coherent diagnostic and prognostic frames is not so likely to occur in an organizational vacuum. Rather, such activity is more likely to flourish in the micromobilization contexts provided by viable SMOs. Why? Because viable SMOs are more likely to provide the raw materials (e.g., alternative ideas) and interactional venues (e.g., meetings) and mechanisms (e.g., talk and debate) conducive to frame articulation and elaboration. We were continually struck by the differences in the SMOs we examined in this regard. The viable ones were more adaptive in that they could strategically develop or modify both diagnostic and prognostic frames in a fashion that facilitated outcome attainment. For example, the DtUH, which had difficulty securing outcomes under conditions that favored other SMOs, was able to attain desired outcomes when their message emphasized the educational consequences of homelessness for children. Had they merely pushed for rights and relief for the homeless in general, it is likely that their appeal would have fallen on deaf ears.

Even though our findings indicate that viable SMOs and articulate and coherent framing activities are mutually facilitative, they also indicate

that working prognostic frames may sometimes develop in the absence of articulate diagnostic frames and viable organizational contexts (pathways 4 and 5). This is not so surprising as it may appear when two considerations are taken into account. First, the grievances associated with some movements, such as the homeless one, are often so palpable because they are rooted in the disruption of constituents everyday routines, that many, and perhaps most, of them have a good sense of some aspect of the problem, thus making the articulation of a diagnostic frame less of a troublesome mobilization issue. However, the specification of some line of action is still likely to be a problematic issue in such contexts, thus requiring the development of a prognostic frame (Snow et al. 1998, pp. 18–19). This takes us to the second consideration: that functional prognostic frames may sometimes be elaborated by a few ardent activists in nonviable organizational contexts, which is exactly what occurred in the cases of the four SMOs associated with pathways 4 and 5 (Boston Union of the Homeless [BUH], DnUH, HU, and HUH). It is our sense, based on the careers of these four SMOs, that prognostic frames developed in these contexts are useful for guiding one or two collective actions but not for mounting a sustained challenge. In order for that to occur, viable organizations that are able to accumulate resources and engage in elaborated and focused framing discussions appear to be requisite conditions.

Whatever the sources or correlates of articulate and coherent diagnostic and prognostic frames, however, it seems clear that they are no less important to movement outcome attainment efforts than organizational structure, tactical considerations, and political context. In fact, our findings indicate that framing activity was the most persistently present condition across all six pathways. As shown in table 6, not only was both diagnostic and prognostic framing present in all three pathways associated with two or more outcomes, but prognostic framing was also present in two of the three pathways associated with only one outcome. In contrast, none of the other conditions were present in more than three of the six pathways. Thus, for the 15 homeless SMOs we examined, it is apparent that articulate and focused framing activity comes more closely than any of the other conditions to constituting a necessary condition for attainment of the outcomes in question. This is not so surprising, since both diagnostic and prognostic framing not only help define and bring into sharp focus grievances, targets of blame, targets of action, and lines of action, but, in doing so, may also help SMOs secure or enhance organizational legitimacy within the organizational fields in which they are embedded. And, in turn, this legitimacy is likely to increase the prospect of securing sympathetic allies and official, city support. Thus, just as Gamson and Meyer (1996) have argued that framing can affect political opportunity, so it appears that organizational viability and legitimacy may be affected by framing

activity as well. In the absence of either type of framing, then, any concerted collective action is unlikely, and this seems particularly clear in the case of prognostic framing.

We believe that the neglect of framing processes in understanding movement outcomes is one of the more glaring oversights in the social movement literature. One need only note some of the century's seemingly most consequential movements—the communist movement, the Civil Rights movement, and the women's movement—to underscore the important role of articulate and coherent diagnostic and prognostic framing. Indeed, our analysis underscores and vitalizes the notion that collective definitions and perceptions matter in social movements. People do not typically respond to abstract notions like political opportunity structure and grievances apart from the creative and inspirational ways in which they are interpreted by movement leaders and activists (Gamson and Meyer 1996). While the framing literature generally has emphasized the role that these factors play in movement mobilization, we have argued here that they also make an important contribution to understanding more fully the factors and dynamics accounting for variation in outcome attainment across movements and contexts.

Summary

Taken together, our findings identify the importance of organizational viability and framing activities for obtaining targeted outcomes by homeless SMOs. When these conditions are present and occur in conjunction with political mediation, the particulars of which affect the types of tactics that are associated with successful outcome attainment, we found that the homeless SMOs are likely to have their greatest impact. While it is an empirical question whether this conjunction of conditions holds for other movements, the findings and analysis suggest that attempts to understand movement outcomes that focus on the ways in which different conditions interact and combine are likely to be more compelling and robust, both theoretically and empirically, than efforts that focus on the conditions specified by a single perspective or that pit one perspective against another. Additionally, while we suspect that different combinations of conditions are likely to be associated with different categories of outcomes for different movements, we would be surprised if the framing processes elaborated here did not figure prominently in successful outcome attainment efforts among most other movements. We say this not only because either articulate and focused diagnostic or prognostic framing, or both, was a necessary condition for securing all four categories of outcomes sought by the 15 SMOs studied, but also because a consideration of framing processes forces analysts to consider not only structural factors, such as orga-

nizational form and political context, but also the various activities movement adherents engage in within the context of SMOs.

REFERENCES

- Amenta, Edwin, Bruce G. Carruthers, and Yvonne Zylan. 1992. "A Hero for the Aged? The Townsend Movement, the Political Mediation Model, and U.S. Old-Age Policy, 1934–1950." *American Journal of Sociology* 98:308–39.
- Amenta, Edwin, Kathleen Dunleavy, and Mary Bernstein. 1994. "Stolen Thunder? Huey Long's 'Share Our Wealth,' Political Mediation, and the Second New Deal." *American Sociological Review* 59:678–702.
- Amenta, Edwin, Robin Tamarelli, and Michael P. Young. 1996. "The Old Folks at Home: Political Mediation Theory and the Impact of the Townsend Movement in California." Unpublished manuscript.
- Amenta, Edwin, and Michael P. Young. 1999. "Making an Impact: Conceptual and Methodological Implications of the Collective Goods Criterion." Pp. 22–41 in *How Social Movements Matter*, edited by Marco Giugni, Doug McAdam, and Charles Tilly. Minneapolis: University of Minnesota Press.
- Babb, Sarah. 1996. "'A True American System of Finance': Frame Resonance in the U.S. Labor Movement, 1866 to 1886." *American Sociological Review* 61: 1033–52.
- Barak, Gregg. 1991. *Gimme Shelter*. New York: Praeger.
- Burawoy, Michael, ed. 1991. *Ethnography Unbound: Power and Resistance in the Modern Metropolis*. Berkeley and Los Angeles: University of California Press.
- Burstein, Paul. 1999. "Social Movements and Public Policy." Pp. 3–21 in *How Social Movements Matter*, edited by Marco Giugni, Doug McAdam, and Charles Tilly. Minneapolis: University of Minnesota Press.
- Burt, Martha R. 1992. *Over the Edge: The Growth of Homelessness in the 1980's*. New York: Russell Sage Foundation.
- Cloward, Richard, and Francis Fox Piven. 1984. "Disruption and Organization: A Reply to Gamson and Schmeidler." *Theory and Society* 13:587–99.
- Cress, Daniel M., and David A. Snow. 1996. "Mobilization at the Margins: Resources, Benefactors, and the Viability of Homeless Social Movement Organizations." *American Sociological Review* 61:1089–109.
- Curtis, Russell L., and Louis A. Zurcher. 1973. "Stable Resources of Protest Movements: The Multi-Organizational Field." *Social Forces* 52:53–61.
- Diani, Mario. 1997. "Social Movements and Social Capital: A Network Perspective on Movement Outcomes." *Mobilization: An International Journal* 2:129–47.
- DiMaggio, Paul J., and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48:147–60.
- Edwards, Bob, and Sam Marullo. 1995. "Organizational Mortality in a Declining Social Movement: The Demise of Peace Movement Organizations in the End of the Cold War Era." *American Sociological Review* 60:908–27.
- Feagin, Joe R., Anthony M. Orum, and Gideon Sjöberg, eds. 1991. *A Case for the Case Study*. Chapel Hill: University of North Carolina Press.
- Frey, R. Scott, Thomas Dietz, and Linda Kalof. 1992. "Characteristics of Successful American Protest Groups: Another Look at Gamson's *Strategy of Social Protest*." *American Journal of Sociology* 98:368–87.
- Gamson, William A. 1990. *The Strategy of Social Protest*, 2d ed. Belmont, Calif.: Wadsworth Publishing.

- . 1992. *Talking Politics*. Cambridge: Cambridge University Press.
- Gamson, William A., and David S. Meyer. 1996. "Framing Political Opportunity." Pp. 275–90 in *Comparative Perspectives on Social Movements: Political Opportunities, Mobilizing Structures, and Cultural Framings*, edited by Doug McAdam, John D. McCarthy, and Mayer N. Zald. New York: Cambridge University Press.
- Gamson, William A., and Emilie Schmeidler. 1984. "Organizing the Poor." *Theory and Society* 13:567–85.
- Giugni, Marco G. 1998. "Was It Worth the Effort? The Outcomes and Consequences of Social Movements." *Annual Review of Sociology* 24:371–93.
- Goldberg, Robert A. 1991. *Grassroots Resistance: Social Movements in Twentieth Century America*. Belmont, Calif.: Wadsworth Publishing.
- Goldstone, Jack A. 1980. "The Weakness of Organization: A New Look at Gamson's *Strategy of Social Protest*." *American Journal of Sociology* 85:1017–42.
- Gusfield, Joseph. 1981. *The Culture of Public Problems: Drinking and Driving and the Symbolic Order*. Chicago: University of Chicago Press.
- Haines, Herbert H. 1984. "Black Radicalization and the Funding of Civil Rights: 1957–1970." *Social Problems* 32:31–43.
- Hutchinson, Richard N., Andrew W. Jones, Nella Van Dyke, Leslie Gates, and Michele Companion. 1997. "Local Mesomobilization: Informal Organization and Networks in the Accomplishment of Collective Action." Unpublished manuscript.
- Issac, Larry, and William R. Kelley. 1981. "Racial Insurgency, the State, and Welfare Expansion: Local and National Level Evidence from the Postwar United States." *American Journal of Sociology* 86:1348–86.
- Jenkins, J. Craig. 1981. "Sociopolitical Movements." Pp. 81–153 in *The Handbook of Political Behavior*, vol. 4. Edited by S. Long. New York: Plenum Press.
- Jenkins, J. Craig, and Craig Eckert. 1986. "Channeling Black Insurgency: Elite Patronage and Professional Social Movement Organization in the Development of the Black Movement." *American Sociological Review* 51:812–29.
- Klandermans, Bert. 1992. "The Social Construction of Protest and Multiorganizational Fields." Pp. 77–103 in *Frontiers in Social Movement Theory*, edited by Aldon D. Morris and Carol M. Mueller. New Haven, Conn.: Yale University Press.
- Marx, Gary, and James Wood. 1975. "Strands of Theory and Research in Collective Behavior." *Annual Review of Sociology* 1:363–428.
- McAdam, Doug. 1982. *Political Process and the Development of Black Insurgency, 1930–1970*. Chicago: University of Chicago Press.
- . 1988. *Freedom Summer*. New York: Oxford University Press.
- McAdam, Doug, John D. McCarthy, and Mayer Zald. 1988. "Social Movements." Pp. 695–737 in *Handbook of Sociology*, edited by Neil Smelser. Beverly Hills, Calif.: Sage Publications.
- McCarthy, John D. 1994. "Activists, Authorities, and Media Framing of Drunk Driving." Pp. 133–67 in *New Social Movements: From Ideology to Identity*, edited by Enrique Larana, Hank Johnston, and Joseph R. Gusfield. Philadelphia: Temple University Press.
- Meyer, David S. 1990. "Peace Movement Demobilization: The Fading of the Nuclear Freeze." Pp. 53–71 in *Peace Action in the Eighties: Social Science Perspectives*, edited by Sam Marullo and John Lofland. New Brunswick, N.J.: Rutgers University Press.
- Minkoff, Debra C. 1993. "The Organization of Survival: Women's and Race-Ethnic Voluntary and Activist Organizations, 1955–1985." *Social Forces* 71:887–908.
- Mirowsky, John, and Catherine E. Ross. 1981. "Protest Group Success: The Impact of Group Characteristics, Social Control, and Context." *Sociological Focus* 14:177–92.
- Morris, Aldon D. 1984. *The Origins of the Civil Rights Movement: Black Communities Organizing for Change*. New York: Free Press.

- Piven, Francis Fox, and Richard A. Cloward. 1977. *Poor People's Movements*. New York: Vintage Books.
- . 1992. "Normalizing Collective Protest." Pp. 301–25 in *Frontiers in Social Movement Theory*, edited by Aldon D. Morris and Carol M. Mueller. New Haven, Conn.: Yale University Press.
- Ragin, Charles C. 1987. *The Comparative Method*. Berkeley and Los Angeles: University of California Press.
- . 1989. "The Logic of the Comparative Method and the Algebra of Logic." *Journal of Quantitative Anthropology* 1:373–98.
- Rosenthal, Rob. 1994. *Homeless in Paradise: A Map of the Terrain*. Philadelphia: Temple University Press.
- Rossi, Peter H. 1989. *Down and Out in America: The Origins of Homelessness*. Chicago: University of Chicago Press.
- Snow, David A., and Leon Anderson. 1993. *Down on Their Luck: A Study of Homeless Street People*. Berkeley and Los Angeles: University of California Press.
- Snow, David A., and Robert D. Benford. 1988. "Ideology, Frame Resonance, and Participant Mobilization." *International Social Movement Research* 1:197–217.
- . 1992. "Master Frames and Cycles of Protest." Pp. 133–55 in *Frontiers in Social Movement Theory*, edited by Aldon D. Morris and Carol M. Mueller. New Haven, Conn.: Yale University Press.
- Snow, David A., Robert D. Benford, and Leon Anderson. 1986. "Fieldwork Roles and Informational Yield: A Comparison of Alternative Settings and Roles." *Urban Life* 14:377–408.
- Snow, David A., Daniel M. Cress, Liam Downey, and Andrew W. Jones. 1998. "Disrupting the 'Quotidian': Reconceptualizing the Relationship between Breakdown and the Emergence of Collective Action." *Mobilization: An International Journal* 3:1–22.
- Snow, David A., Burke Rochford, Steve Worden, and Robert D. Benford. 1986. "Frame Alignment Processes, Micromobilization, and Movement Participation." *American Sociological Review* 51:464–81.
- Steadly, Homer R., and John W. Foley. 1979. "The Success of Protest Groups: Multi-Variate Analysis." *Social Science Research* 8:1–15.
- Tarrow, Sidney. 1992. "Mentalities, Political Cultures, and Collective Action Frames: Constructing Meaning through Action." Pp. 174–202 in *Frontiers in Social Movement Research*, edited by Aldon Morris and Carol McClurg Mueller. New Haven, Conn.: Yale University Press.
- . 1994. *Power in Movement: Social Movements, Collective Action, and Politics*. New York: Cambridge University Press.
- Voss, Kim. 1996. "The Collapse of a Social Movement: The Interplay of Mobilizing Structures, Framing, and Political Opportunities in the Knights of Labor." Pp. 227–58 in *Comparative Perspectives on Social Movements*, edited by Doug McAdam, John D. McCarthy, and Mayer N. Zald. New York: Cambridge University Press.
- Wagner, David. 1993. *Checkerboard Square: Culture and Resistance in a Homeless Community*. Boulder, Colo.: Westview Press.
- Wright, James D. 1989. *Address Unknown: The Homeless in America*. New York: Aldine de Gruyter.
- Wright, Talmadge. 1997. *Out of Place: Homeless Mobilizations, Subcities, and Contested Landscapes*. Albany: State University of New York Press.
- Zald, Mayer N., and Roberta Ash. 1966. "Social Movement Organizations: Growth, Decay, and Change." *Social Forces* 44:327–41.
- Zald, Mayer N., and Bert Useem. 1987. "Movement and Countermovement Interaction: Mobilization, Tactics, and State Involvement." Pp. 247–72 in *Social Movements in an Organizational Society*, edited by Mayer N. Zald and John D. McCarthy. New Brunswick, N.J.: Transaction Books.

The Emergence of Generalized Exchange¹

Nobuyuki Takahashi
University of Arizona

The existence of generalized exchange characterized by unilateral resource giving has been a puzzle when we assume rational actors, because free riding can occur. This article first identifies pure-generalized exchange in which each actor gives resources to the recipient(s) of his choice. Then, it proposes the fairness-based selective-giving strategy. An actor adopting this strategy selects a recipient whose behaviors satisfy her criterion of fairness, provided perfect information is given. The results of evolutionary simulation show that pure-generalized exchange can emerge among egoists without collective norms, even in societies in which individuals have information only about their immediate neighbors.

INTRODUCTION

Generalized exchange has been one of the central topics in the classical social exchange literature both in sociology and anthropology (e.g., Befu 1977, 1980; Blau 1964; Ekeh 1974; Heath 1976; Lévi-Strauss 1949; Malinowski 1922; Mauss [1925] 1954; Sahlins 1972). Generalized exchange has been conceived as one of the mechanisms that enhances social solidarity. However, little empirical work has been done on generalized exchange (Emerson 1976, 1981; Gillmore 1987; Heath 1976). As a result, a main question remains unanswered: Why does generalized exchange emerge, and how is it maintained? This question addresses the mechanism that can generate and maintain generalized exchange systems. Generalized exchange is typically characterized by unilateral resource giving (Molm and Cook 1995) because one's giving is reciprocated not by the recipient, but by the third party. Thus, from the viewpoint of social exchange, rational choice, or evolutionary theory, the existence of generalized exchange is a puzzle because any member of the exchange system can free ride. There is no guarantee of reciprocity. Therefore, previous researchers have ex-

¹ Research presented in this article was partly supported by a scientific research grant from the Japanese Ministry of Education. I am grateful for the helpful comments of Sun-Ki Chai, Michael Hechter, Alan Miller, Linda Molm, and James Ranger-Moore. Direct correspondence to Nobuyuki Takahashi, Department of Sociology, College of Social and Behavioral Sciences, P.O. Box 210027, Tucson, Arizona 85721-0027.

plained generalized exchange as the result of altruism (Sahlins 1972; Takagi 1994, 1996) or collective norms (Ekeh 1974; Lévi-Strauss 1949). Recently, however, several new attempts to explain the emergence of generalized exchange have been made (e.g., Bearman 1997; Takagi 1994, 1996; Takahashi and Yamagishi 1996, 1999; Ziegler 1990). However, these solutions are typically limited to forms of generalized exchange that have a fixed network structure. Based on this research, especially Takahashi and Yamagishi (1996), this article develops a more plausible and general solution to the problem of the emergence of generalized exchange that holds even when none of the previous requirements are satisfied.

What Is Generalized Exchange?

Social exchange theory has roots in multiple disciplines. These include sociology (e.g., Homans [1961] 1974; Blau 1964; Emerson 1972*a*, 1972*b*; Heath 1976), psychology (e.g., Thibaut and Kelley 1959), and anthropology (e.g., Lévi-Strauss 1949; Ekeh 1974; Sahlins 1972). Although different versions of social exchange theory use different terminologies, most share a key distinction between restricted (or direct) exchange and generalized (or indirect) exchange (e.g., Bearman 1997; Befu 1980; Blau 1964; Ekeh 1974; Emerson 1976, 1981; Heath 1976; Gillmore 1987; Lévi-Strauss 1949; Molm and Cook 1995; Sahlins 1972; Takagi 1996; Takahashi and Yamagishi 1996, 1999; Yamagishi and Cook 1993).

In restricted exchange, two actors exchange resources with each other. In other words, the resources that one actor gives are directly contingent on the resources that the other gives in return. If *A* gives to *B*, *B* is the person who would reciprocate to *A*. This type of exchange is very common. Examples include exchanges between teachers and students, economic transactions, employers and employees, and so on. Most of the social exchange network research that has emerged since the 1980s in sociology focuses only on restricted exchange (e.g., Bienenstock and Bonacich 1992; Cook and Emerson 1978; Cook, Emerson, Gillmore, and Yamagishi 1983; Friedkin 1992; Markovsky, Willer, and Patton 1988; Yamagishi, Gillmore, and Cook 1988).

In contrast to restricted exchanges, which occur between two actors, generalized exchange inherently involves more than two people. In generalized exchange, there is no one-to-one correspondence between what two actors directly give to and receive from each other. *A*'s giving to *B* is not reciprocated by *B*'s giving to *A*, but by *C*'s giving to *A*, where *C* is a third party. Thus, reciprocation is indirect, not mutual. *A* gives help to *B* when *B* is in need, and at cost to *A*. If and when *A* needs help, and other actors (*C*, *D*, . . . , *N*) are around, one of them may provide it—again, with no

assurance of return (Emerson 1981). "If I see burglars in my neighbor's house, I have the duty of doing something about it (e.g., calling the police), not because I expect any reciprocation—of whatever type from my unfortunate neighbor—but perhaps because I expect any neighbor of mine to do the same thing if he sees burglars in my own house" (Ekeh 1974, p. 206).

At first sight, this does not appear to be an exchange at all. However, each actor provides resources at some time and eventually receives some benefit in return—not from the same actor, but from a different actor. In this sense, exchange theorists have traditionally considered generalized exchange as one type of exchange. The classic examples of generalized exchange are the Kula ring (e.g., Malinowski 1922; Ziegler 1990) and matrilineal cross-cousin marriage (e.g., Lévi-Strauss 1949; Bearman 1997). These forms are special cases of generalized exchange because in each there is a fixed network structure of exchange, that is, a chain. Generalized exchange is present, at least in a rudimentary form, in many aspects of social life (Emerson 1981; Heath 1976; Gillmore 1987). Examples include revolving credit associations, duplicate bridge games in which the players cycle through hosts, reviewers of journal articles, helping a stranded driver on a mountain road, the anonymous donation of blood, giving shower gifts, villagers going from household to household helping in harvesting, and so on (e.g., Bearman 1997; Befu 1977; Emerson 1981; Molm and Cook 1995; Takagi 1996).

One may have noticed that the existence of such a system is problematic. People who engage in unilateral resource giving in generalized exchange systems do not expect nor receive a direct return from the recipient, although they may expect a return from someone else in the future (e.g., Befu 1977; Ekeh 1974; Sahlins 1972; Yamagishi and Cook 1993). Therefore, every member of a generalized exchange system can (but need not) receive resources if everyone gives his or her resources to someone else. Thus, forming a generalized exchange system is very risky because unilateral resource giving is an invitation to exploitation (e.g., Bearman 1997; Gillmore 1987; Yamagishi and Cook 1993). This feature of generalized exchange coincides with the problem that is prevalent in another research area: the free rider problem of social dilemmas (e.g., Takagi 1996; Yamagishi and Cook 1993; Ziegler 1990). In other words, generalized exchange has the characteristic of social dilemmas (e.g., Molm 1994; Takagi 1996; Yamagishi and Cook 1993; Ziegler 1990). Rational self-interested members will be better off if they do not give resources to others. However, members who think that others will not give are unlikely to give away their own resources, and generalized exchange may never be established. Thus, the establishment and maintenance of a generalized exchange system requires the solution of a social dilemma.

PREVIOUS RESEARCH TRADITIONS

Generalized exchange is characterized by unilateral resource giving. But how and why does such an exchange system exist? In this section, I first briefly review what has been argued by the classic researchers. Then, I review several works that are directly related to this article.

The first type of explanation is that people give resources unilaterally because they have an altruistic motivation. Although there are debates about the definition of altruism, here it simply means that an actor's behaviors are based on concern for others' well-being.² Sahlin (1972), for example, characterized generalized exchange in terms of pure gifts with no obligation to repay. Thus, the result of unilateral resource giving or altruism is the generalized exchange system (Takagi 1994, 1996). However, the utility of this explanation is limited, because it raises another question: Why do people have such a motivation? Some researchers answered socialization to this question (e.g., Nye 1979), but this raises another question: Why is there such a socialization pressure in a society? These answers may hold one or the other—not both.

The second type of explanation requires a collective norm, the unilateral reciprocity principle. Once an actor receives resources, she is obligated to return to someone else in the future. In other words, free riding is a violation of this norm. Therefore, the social dilemma should not exist in a society characterized by such a collective norm of unilateral reciprocity. And, generalized exchange generates a morality characterized by credit mentality (e.g., Ekeh 1974; Lévi-Strauss 1949). Obviously, such an argument is the functional-collectivist one. There is no explanation of where such a norm comes from. Moreover, generalized exchange should not exist in the absence of such a norm.

The third type of explanation employs a rational choice/game theoretic framework. People give resources unilaterally when this behavior is beneficial to them. This answer assumes rational actors and considers unilateral resource giving as an instrumental behavior in order to gain other benefits. This is one of the most discussed ideas for solving social dilemmas or free rider problems.³ One of the most well-known answers is the use of selective incentives (e.g., Olson 1965). Olson argued that we have to change the incentive structure in order to solve social dilemmas. For example, a strong organization can sanction members so that it is more beneficial for each member to contribute to the group goal. However, this solution implicitly assumes that actors' behavior can be monitored so that

² See Sesardic (1995) for a recent review.

³ For reviews, see Dawes (1980), Messick and Brewer (1983), Orbell and Dawes (1981), Stroebe and Frey (1982), Yamagishi (1989a, 1995).

selective incentives can be efficiently applied, and assumes that the group has the necessary means to provide the incentives in the first place (Gillmore 1987).

In order to overcome these limitations, several new solutions to social dilemmas have been proposed since the 1980s. These solutions suggest that people sometimes change the incentive structure voluntarily so that free riding is impossible. In this article, I adopt this approach and apply it to the generalized exchange problem. In the next section, I selectively review these recent works to develop the logic of the new model that I propose in this article.

SOLUTIONS OF THE FREE RIDER PROBLEM

These solutions have several common features. First, they assume rational actors and no central authority. Second, the solutions employ strategies in which actors' behaviors are contingent on others' behaviors. Third, they consider the free rider problem to be solved when certain strategies that involve giving resources to others can obtain higher profit than other strategies. If such a strategy is possible in a situation, rational actors should adopt this strategy, and eventually there should be no free riders.

There are three distinct solutions to social dilemmas that are directly relevant to this article. They are (1) tit-for-tat (TFT) in iterated prisoner's dilemmas, (2) out-for-tat (OFT) in prisoner's dilemma networks, and (3) downward tit-for-tat in network-generalized exchange. I will use the underlying logic and principles in these three solutions in order to develop the solution to the problem of the emergence of generalized exchange. For other solutions in sociology, not directly relevant to this article, see Macy's (1990, 1991a, 1991b, 1993a, 1993b) and Heckathorn's (1990, 1993, 1996) series of work.

The Tit-for-Tat Strategy in Iterated Prisoner's Dilemmas

Tit-for-tat (TFT) in the social dilemmas literature, balanced reciprocity in social exchange theory (Sahlins 1972), and reciprocal altruism in biology are based on similar ideas. Given that people are embedded in long-lasting relationships, resource-giving behavior can be profitable. Suppose *A* can give some valuable resources to *B*. If *A* gives to *B*, *B* will give something to *A* in the future. However, if *A* does not give to *B*, *B* will not give either. In research on the prisoner's dilemma (PD), such a behavior is called a tit-for-tat strategy. It is defined as follows. In an iterated PD, an actor adopting TFT starts with cooperation. In subsequent rounds, an actor adopting TFT cooperates if and only if the partner cooperated in the previous round. Otherwise, he defects. The effectiveness of the TFT strategy

has been widely shown both theoretically and empirically (see Axelrod 1980*a*, 1980*b*, 1984; Oskamp 1971; Wilson 1971). In the game theoretic literature, TFT has been one of the best strategies in iterated PDs. Many experiments have shown that TFT produces mutual cooperation. However, it works only in an iterated PD that corresponds to restricted—not generalized—exchange. An actor adopting the TFT strategy does not necessarily give resources to others when there are no long-lasting dyads. Therefore, we need another mechanism.

Out-for-Tat Strategy in Prisoner's Dilemma Networks

The second solution to the free rider problem is to use the out-for-tat (OFT) strategy in prisoner's dilemma networks (Hayashi, Jin, and Yamagishi 1993; Yamagishi, Hayashi, and Jin 1994).⁴ A prisoner's dilemma network is a situation where (1) every member of a group chooses a partner, and (2) a PD game is played by members who have chosen each other. In this situation, computer simulation tournaments showed that the OFT strategy is overall the most effective strategy among the known strategies and brings mutual cooperation.⁵ The definition of OFT is this. (1) An actor adopting the OFT strategy always cooperates, (2) she sticks with the current partner insofar as the partner cooperates, and (3) she deserts and seeks out a new partner as soon as the partner defects. Because actors adopting the OFT strategy do not choose defectors, eventually they commit to each other. These pairs of actors adopting the OFT strategy can attain mutual cooperation.⁶ Contrary to this, the possible partners for defectors are other defectors. Thus, the OFT strategy can produce higher profit than defectors receive. Therefore, in prisoner's dilemma networks, a PD is voluntarily solved by the OFT strategy. Now, we can explain resource-giving behavior in any dyad by either TFT or OFT. However, PD networks still correspond to restricted exchange because the PD involves bilateral resource giving (Yamagishi and Cook 1993). We need a strategy that can solve the free rider problem when there is only unilateral resource giving.

⁴ For the review of the recent development of the "selective play" paradigm, see Yamagishi and Hayashi (1996).

⁵ For the results of the computer tournament and the detailed discussion of the effectiveness of the OFT strategy, see Yamagishi, Hayashi, and Jin (1994), Hayashi (1995), Yamagishi and Hayashi (1996), Hayashi and Yamagishi (1998).

⁶ The OFT strategy is the best only under certain conditions: small number of actors adopting the random strategy and no opportunity costs (e.g., Yamagishi, Hayashi, and Jin 1994; Yamagishi and Hayashi 1996; Hayashi and Yamagishi 1998).

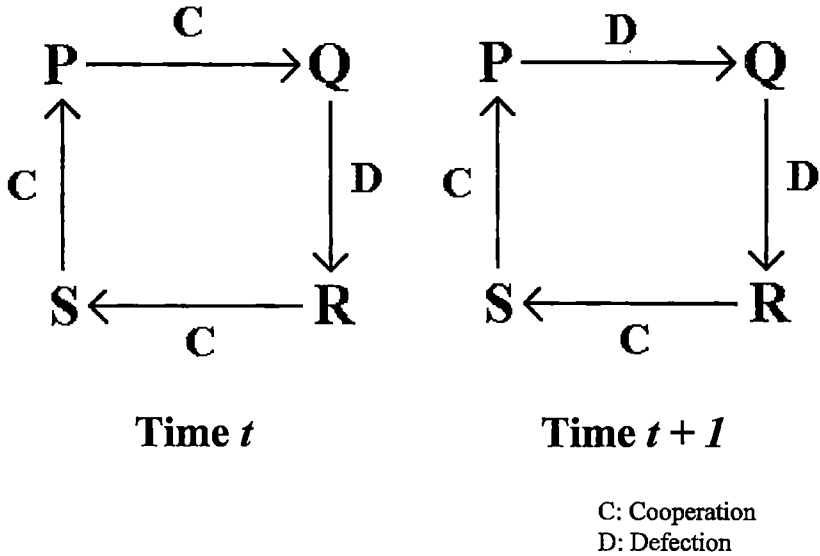


FIG. 1.—Downward TFT in network-generalized exchange

Downward Tit-for-Tat in Network-Generalized Exchange

Only a few empirical studies of generalized exchange have been conducted since the classical studies on the Kula rings and matrilineal cross-cousin marriage in anthropology (e.g., Bearman 1997; Gillmore 1987; Takahashi and Yamagishi 1999; Uehara 1990; Yamagishi and Cook 1993). Among them, Yamagishi and Cook (1993) is the most relevant here. They distinguished two forms of generalized exchange: “group-generalized” exchange and “network-generalized” exchange.⁷ In the first type, group members pool their resources and then receive the benefits that are generated by pooling. In the second structure, which is the focus here, (1) each actor gives resources to an actor in the network who does not return resources directly to that actor, and (2) instead, the giver receives resources from some other actor in the network. An example of network-generalized exchange is shown in figure 1. The structure of exchange can be much more complex.

Yamagishi and Cook’s experiment showed that people give more resources in network-generalized exchange than in group-generalized exchange. They explained this difference by the difference of the effectiveness of strategies. In group-generalized exchange, reciprocal strategies

⁷ In Ekeh’s (1974) words, this is “chain-generalized exchange.”

(e.g., TFT) are not usually effective because one actor's behavior affects all the others (for more complete discussions of the effectiveness of reciprocal strategies in N -person situations, see Dawes [1980] and Yamagishi [1989b]). In network-generalized exchange, however, the actor who unilaterally gives resources has total control over the recipient's outcomes. Therefore, if the giver notices that the recipient is not giving to the recipient's recipient, he can punish the recipient by withholding resources until the recipient gives her resources to the recipient's recipient.

In another important study in a different field, biologists Boyd and Richerson (1989) tried to explain prosocial behavior by indirect reciprocity. Although most of the research in biology has focused on cooperation in dyads, they focused on the n -person relationship within a group. Particularly, they focused on the chain of helping, which corresponds to network-generalized exchange in Yamagishi and Cook (1993). According to Boyd and Richerson (1989), in network-generalized exchange, each actor can potentially use two types of unilateral-TFT strategies. These strategies in generalized exchange settings are unilateral because the giver can control the recipient's outcome but not vice versa: there is no way for the recipient to reward or punish the giver. What Boyd and Richerson (1989) proposed are upward-TFT and downward-TFT strategies. An actor adopting upward TFT gives resources to a designated recipient if and only if she received resources from a designated giver. In figure 1, Q gives resources to R at time $t + 1$ if and only if Q received resources from P at time t . Contrary to this rule, an actor using downward-TFT gives resources to a designated recipient if and only if the recipient gave to his own designated recipient in the previous trial. In figure 1, actor P gives her resources to Q at time $t + 1$ if and only if Q gives her resources to R at time t . Based on evolutionary biology and mathematical analysis, Boyd and Richerson (1989) argued that a network-generalized exchange system might be sustained if actors adopted the downward-TFT strategy.⁸

These two studies agree with each other in the conclusion that when network-generalized exchange emerges, it *may* be the result of a downward-TFT strategy. When people adopt this strategy, unilateral resource giving can be profitable. Free riding is not possible. There is at least one serious limitation in this solution, however. For generalized exchange to emerge, a particular fixed network structure must last for a long time.⁹

⁸ Strictly speaking, Boyd and Richerson (1989) only argued that generalized exchange is more likely when each actor adopts downward TFT than when each actor adopts upward TFT. They admitted that even if actors adopt downward TFT, generalized exchange may be very unrobust.

⁹ Another limitation is that these solutions might work only in extremely small groups. One reason is that network-generalized exchange is vulnerable to only one "hard-core" defector. Boyd and Richerson (1989) admit this possibility. Yamagishi and Cook

The network must be fixed and cannot be flexible. Otherwise, downward TFT is useless. However, many of the generalized exchange systems in the real world are not a simple closed chain of resource giving as in figure 1. For example, we do not always help the same stranded driver. Kula rings are more complex than a simple chain (Ziegler 1990). Empirically, there are certain deviations from the pattern of the chain of marriage (Bearman 1997). Thus, we need a more general solution to the free rider problem when there is no fixed network structure of generalized exchange. I present such a solution in the next section.

NEW MODEL: FAIRNESS-BASED SELECTIVE GIVING IN PURE-GENERALIZED EXCHANGE

In network-generalized exchange, there is a fixed network structure. Contrary to this, we can consider generalized exchange where there is no fixed structure. In Ekeh's (1974) example of the witness of a burglary, there is no fixed network of give and receive in the community. In this article, I call this situation *pure-generalized exchange* (Takahashi and Yamagishi 1995, 1996). Pure generalized is the most general, flexible, and least restricted type of exchange. In pure-generalized exchange, each actor gives resources to a recipient(s) that he chooses unilaterally. An example is shown in figure 2. Pure-generalized exchange is network-generalized exchange with a choice of recipients.

As Takahashi and Yamagishi (1995, 1996) have shown, the downward-OFT strategy can solve the free rider problem in a pure-generalized exchange system. This strategy is a synthesis of the OFT strategy and the downward-TFT strategy. We can consider this strategy as a variation of the OFT strategy where there is no direct reciprocity. Alternatively, we can consider this as a variation of the downward-TFT strategy where there is no fixed "chain" of exchange. However, there is an additional feature in this strategy. An actor using the downward-OFT strategy has to decide her behavior based on the recipient's behavior not to herself but to the third party. Takahashi and Yamagishi (1995, 1996) assumed that an actor compares her own behavior to the recipient's behavior and decides what to do to the recipient based on certain criterion. They called this decision-making mechanism "a sense of fairness." Thus, more precisely, an actor using the downward-OFT strategy (1) keeps giving resources to his recipients insofar as the recipient's behavior pattern of giv-

(1993) also admit this possibility, but they did not find a negative effect of group size on cooperation rate when they compared four-actor and eight-actor networks. The other reason is that it is necessary for each actor to know his designated recipient's behavior in order to use the downward-TFT strategy.

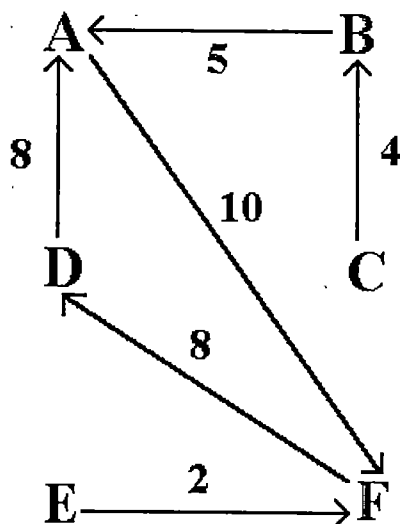


FIG. 2.—Selective giving in pure-generalized exchange. The number of each arrow represents how much each actor gives to a recipient. This giving amount is completely determined by each actor's giving gene. Each actor's tolerance gene is assumed to be 0 in figure 2.

ing meets the actor's criterion of fairness, and (2) deserts the recipient if the recipient does not meet his criterion and selects a new recipient randomly. An advantage of this solution is that an actor adopting the downward-OFT strategy requires information only about his recipient's behavior; therefore, the solution applies to a group of any size. However, it has one serious limitation. For pure-generalized exchange to emerge, the fairness criterion must be the same among all members of a group. In other words, a collective sense of fairness is necessary.

In this article, I propose a new model that eases this limitation, based on the fairness-based selective-giving strategy. I define it as follows: (1) An actor adopting the fairness-based selective-giving strategy gives her resources to a recipient whom she selects. (2) The actor selects a recipient whose giving behavior satisfies her criterion based on information from the previous behaviors. (3) The criterion is determined by the sense of fairness of this actor comparing the giving amount of this actor and that of the potential recipients.

This strategy has several features. First, it has the features of the unilateral conditional strategies. Because there is no direct exchange of resources in generalized exchange, any strategy must work unilaterally. There is no way for potential recipients to affect the giver's outcome.

Second, I assume that each actor has his own criterion of fairness and

determines to whom he gives based on this sense of fairness. What is called "fairness" here has several characteristics. First, it refers to the comparison between how much the potential recipients give and how much the giver gives. In other words, it is the giver's evaluation of the ratio between how much the potential recipients receive from the giver (i.e., output) and how much the potential recipients give to others (i.e., input). In this sense, this is the same as what has been proposed in various fairness/equity/distributive justice theories in social psychology (e.g., Adams 1965; Homans 1974; Walster, Walster, and Berscheid 1978). All of them are concerned with the ratio of input and output. However, there is a big difference as well. While fairness involves the concern for ego's outcome in the fairness/equity/distributive justice research and mathematical models of justice in sociology (e.g., Jasso 1980; Markovsky 1985), in this pure-generalized exchange setting, "fairness" has nothing to do with ego's outcome. Each actor chooses a recipient based on his sense of fairness. Each actor can either give to a very generous actor or a very stingy actor. However, this choice of recipient makes no difference to the amount the giver receives, because there is no way for the recipient to affect the giver's outcome. Although a sense of "fairness" that is unrelated to one's own outcome seems strange at first sight, we certainly have this sense of "fairness." We like to help those who help others more than those who do not help others. If you encounter a car accident involving two injured people trapped in the cars, one who has worked hard to make the community better and the other who has not participated in any community activities, which person would you try to help first? Probably most people would help the first person first. The literature on helping/prosocial behavior has shown that whether the recipient deserves the help or not has an impact on helping (see, e.g., Batson [1987] and Piliavin and Charng [1990] for the review).

Third, because I assume that there can be a variety of standards of fairness in one society, this assumption is different from assuming that fairness is a collective norm (e.g., "univocal norm" in Ekeh [1974]). The criterion is not determined collectively by the society as a whole but by each individual. A recipient who is considered to be fair by one actor may be considered unfair by another actor.¹⁰

Fourth, an actor adopting this strategy punishes stingy actors (e.g., non-givers), not by imposing negative sanctions, but by not giving. Each actor's criterion of fairness determines who is stingy and who is not. This

¹⁰ As one of the reviewers pointed out, this formulation still imposes a uniform norm of fairness because everyone has some criteria and everyone has the idea that giving more is better. The point I am making here is that everyone does not have the same criterion of fairness, although everyone does have, more or less, a sense of fairness.

characteristic has an important theoretical implication. As I discussed above, one of the main solutions to the free rider problem is to impose a penalty on defectors or to give a reward to cooperators. These are called selective incentives (Olson 1965). However, this solution implies that there is an agency to administer selective incentives. Thus, this solution introduces another problem: Who pays the cost of creating such an agency, and who pays the cost of sanctioning? Whether either a positive sanction (reward) or negative sanction (penalty) is administered by a central authority or by individuals, it produces another dilemma, called a "second order dilemma" (e.g., Heckathorn 1988, 1989; Oliver 1980; Yamagishi 1986*a*, 1986*b*). However, an actor adopting the fairness-based selective-giving strategy does not have to pay any cost to desert stingy others. She can simply choose another actor who gives much more. Thus, it does not produce a second-order dilemma.¹¹ This is the most important feature of the "selective play paradigm" where players are endowed with options for leaving the current relation and forming a new relation (e.g., Orbell and Dawes 1991; Hayashi and Yamagishi 1998; Yamagishi and Hayashi 1996). Recent studies have shown the effectiveness of this new paradigm (e.g., Orbell and Dawes 1991, 1993; Schuessler 1989; Yamagishi, Hayashi, and Jin 1994; Vanberg and Congleton 1992; see also Yamagishi and Hayashi [1996] and Hayashi and Yamagishi [1998] for more elaborated discussions on selective play). Moreover, because this sanctioning reduces the target's profit only, it does not produce a domino effect, or a "conflict spiral" (Lawler 1986). Therefore, not giving is the ideal sanction. Based on this theoretical argument, I assume that sanctioning by not giving would also be effective in this new model.

Finally, in applying this model, I assume that each actor knows all of the others' behaviors. This assumption means that this new model can explain generalized exchange only in a group in which everybody knows everybody else. It cannot really explain other types of generalized exchange, such as helping a stranded driver on a mountain road when the driver is a stranger. I will return to this point in the second simulation and try to loosen this restriction as much as possible.

COMPUTER SIMULATIONS

Evolutionary Simulation

In this study, several computer simulations were conducted. These are "evolutionary" simulations, and the basic structure of these simulations is

¹¹ When there is certain cost to withdrawing resources from the current recipient and seeking a new recipient, sanctioning by not giving produces a second-order dilemma.

based on Axelrod (1986). An evolutionary approach is based on the principle that what works well for an actor is more likely to be used again, while what turns out poorly is more likely to be discarded. Axelrod (1986) argued that there could be several different interpretations of this principle. One is a purely biological mechanism that the more effective individuals are more likely to survive and reproduce. The second is the principle of reinforcement in learning theory that the actors learn by trial and error, keeping effective strategies and altering ones that turn out poorly.¹² In this article, I adopt technical terms of the first interpretation (i.e., gene, natural selection, generation, and mutation) because it is the easiest way to illustrate the content of the simulations based on this evolutionary approach. However, these terms do not have substantive theoretical meaning in this article; I use them only as tools to explain evolutionary simulation.

In this study, I created an imaginary society in a computer that consists of members who have "gene(s)" whose values determine certain behaviors of their bearers. I then examine how this society evolves (e.g., how the values of giving genes in a society change) over generations. The result of evolution is represented by the distribution of the value of gene(s) in a society. The actors who are the members of a society in the final generation are the actors who have gene(s) that have received relatively higher profit. The actors who received lower profit cannot survive, and their gene(s) disappear. The purpose of this simulation is to make clear logically which value of gene(s) brings the maximum profit to each actor in a certain situation.¹³

Basic Flowchart

One simulation consists of many replications, one replication consists of many generations, and one generation consists of several trials. In each trial, the computer gives each actor 10 points. Each actor has two genes, and each decides how much and to whom to give her resources.

In the first generation of each replication, each actor's two genes, a giving gene and a tolerance gene, are determined randomly. The giving gene determines how much an actor gives to his recipient. The tolerance gene determines his potential recipients. In this simulation, each actor gives to only one recipient. The giving gene varies from 0 to 10 points. It determines how much an actor gives to a recipient. The actor whose giving

¹² A third one is purposive learning in that the actors observe each other, and those with poor performance tend to imitate the strategies of those they see doing better.

¹³ This simulation is not a substitute for an experiment; instead, it is a substitute for a mathematical analysis.

gene is 0 gives nothing to his recipient in each trial. The actor whose giving gene is 10 gives all of the resources from the computer. The amount of the resources that each actor saves in each trial is added to his total profit in the generation. Thus, the higher the giving gene is, the more an actor loses. However, the amount of the resources that each actor receives from someone else is doubled and added to his total profit.¹⁴ Because resource giving is unilateral, for each actor there is no guarantee of receiving resources from someone else.

Each actor adopts the fairness-based selective-giving strategy and selects a recipient based on her own criterion. This criterion, M , is based on the sense of fairness of each actor and is determined by the giving gene and the tolerance gene:

$$M = \text{an actor's giving gene} \times \text{the actor's tolerance gene.}$$

This tolerance gene is introduced here to make the criterion of fairness variable across actors. If two people who have the same value of the giving gene choose the recipient based on the same criterion of fairness, it means that we have a uniform norm of fairness. However, because of the tolerance gene, these two people may still behave differently. For example, let us suppose that a person has a high giving gene. In the extreme case such as a saint, she would also have a high tolerance gene. She would give a lot and would not discriminate between generous people and stingy people as recipients. However, most people who are generous would prefer to give to other generous people. And, who is generous and who is not from the viewpoint of the giver is determined by the combination of these genes. For example, a very discriminating giver might feel that only those who give more than twice of this giver's giving amount are generous and deserved to receive resources from him. An indiscriminating giver (i.e., closer to a saint) might feel that all the people who give more than half of her giving amount are generous and deserve to be given to by her.

Because this simulation assumes perfect information, each actor knows how many points all of the other actors gave to someone else in the previous trial. Each actor makes a list of actors who gave an amount equal to or more than his criterion. Then, each actor randomly chooses only one recipient from the list. If there is no one who gave more than M in the previous trial, the actor seeks "the second best," choosing the actor who gave the maximum (but still below M) points in the group in the last trial

¹⁴ Following the standard practice in social dilemmas research, I assume that the value of resources received is twice as high as that of one's own resources. However, how many times the value of the given resources to the recipient would not have a big impact on the result, as long as the multiplying factor is larger than 1, which is the necessary condition for social exchange (Takagi 1996).

as a recipient. If the maximum actors are more than 1 in this case, this giver chooses a recipient randomly.

Let me demonstrate how this works in figure 2. In order to simplify this demonstration, assume that the tolerance genes of all of the actors in figure 2 are 1.0 (but in the actual simulations, it varies). First, let us consider *A*. The value of *A*'s giving gene is 10, and the value of *A*'s tolerance gene is 1.0. Therefore, *M* is 10 for *A*. Because none of the others satisfies *M* in figure 2, *A* must choose the second best. *F* and *D*, who give 8 points each, are the second best. Of the two, *A* randomly chooses *F* as his recipient. Next, consider *D*. The actors who satisfy *D*'s criterion are *A* and *F*, and *D* selects *A* as a recipient in this trial. Similarly, the actors who satisfy *C*'s criterion are *A*, *B*, *D*, and *F*, and *C* selects *B* in this figure. However, unfortunately, *C* receives points from no one in this trial. The only potential giver to *C* is *E*, because *E* is the only actor who gives less than *C*. However, *E* gives to *F* in this trial.

One generation consists of 10 trials. After the tenth trial, "natural selection" and "mutation" determine the members of a society in the next generation. In natural selection, each actor's cumulative profit in the generation is compared to others' total profit. An actor whose performance was poor is replaced by an actor whose performance was successful.¹⁵ After that, mutation changes the value of each actor's genes by a small probability.

Simulation 1

To see whether generalized exchange can emerge in a society that consists of various members is the purpose of this first simulation. Table 1 shows the parameters. In this simulation, there is no independent variable. The dependent variable is the mean of the giving gene of each group in the final 10 generations. I conducted two sets of simulations. In simulation 1-1, the initial value of the giving gene was determined randomly. However, even if pure-generalized exchange could emerge in this simulation, this might not be enough to conclude that it can emerge in various societies. In the first generation, there are some members whose giving genes are high. Therefore, the emergence of generalized exchange (if possible) might be the effect of preexisting altruistic actors (i.e., the actors whose giving genes are high). To eliminate this possibility, I conducted the second set of simulations. In simulation 1-2, all giving genes are initially set

¹⁵ The actors whose cumulative profits were less than the criterion (i.e., average profit in the group-standard deviation) leave no offspring. The actors whose cumulative profits were more than the criterion leave two offspring. The actors whose cumulative profits were close to the average leave one offspring. For simplicity, I adjusted the group size to be constant.

TABLE 1
PARAMETERS OF SIMULATION 1

Parameter Names	Values
Group size	20
Number of replications	50
Number of generations	1,000
Number of trials per generation	10
Number of resources that each actor is given in each trial	10
Value of resources that each actor receives from others	2
Giving gene	from 0 to 10
Tolerance gene	from .1 to 2.0
Mutation rate05 (5%)

to 0. This means that no actor gives to anybody in the first generation. The initial value of the tolerance gene was determined randomly in both simulations.

To obtain my measure of the dependent variable, generalized exchange, I first calculated the mean of the giving gene among the group in the last 10 trials, and I then calculated the mean of 50 replications as the mean of each condition. The results are very clear. The level of generalized exchange during the last 10 trials was 9.30 in simulation 1-1, and 9.47 in simulation 1-2. Standard deviations were 1.25 and 0.65, respectively. Thus, there is no substantial difference between the results of these two sets of simulations. Pure-generalized exchange emerged from a society with no altruists, that is, a society in which no one gave to anyone initially. The mean of the tolerance gene during the last 10 trials was 1.00 in both simulations. Standard deviations were 0.22 and 0.27, respectively. Although the value of the tolerance gene varied across members, generalized exchange nonetheless emerged.

Figure 3 shows one example of the history of evolution. This graph shows that generalized exchange is robust. Although the average giving gene declined several times, it soon recovered. This means that a mutant whose giving gene is more than 1 can get more profit than others whose giving genes are 0.¹⁶ Therefore, generalized exchange can emerge even among egoists. In contrast, the mean of the tolerance gene fluctuated. Under the parameters of this simulation, the value of the tolerance gene does

¹⁶ If only one mutant, whose giving gene is more than 0, emerges in a society where no one gives to anyone, he cannot profit. This mutant simply loses his resources. However, when more than one mutant is born in the same generation, these mutants might survive and eventually dominate others whose giving genes are 0.

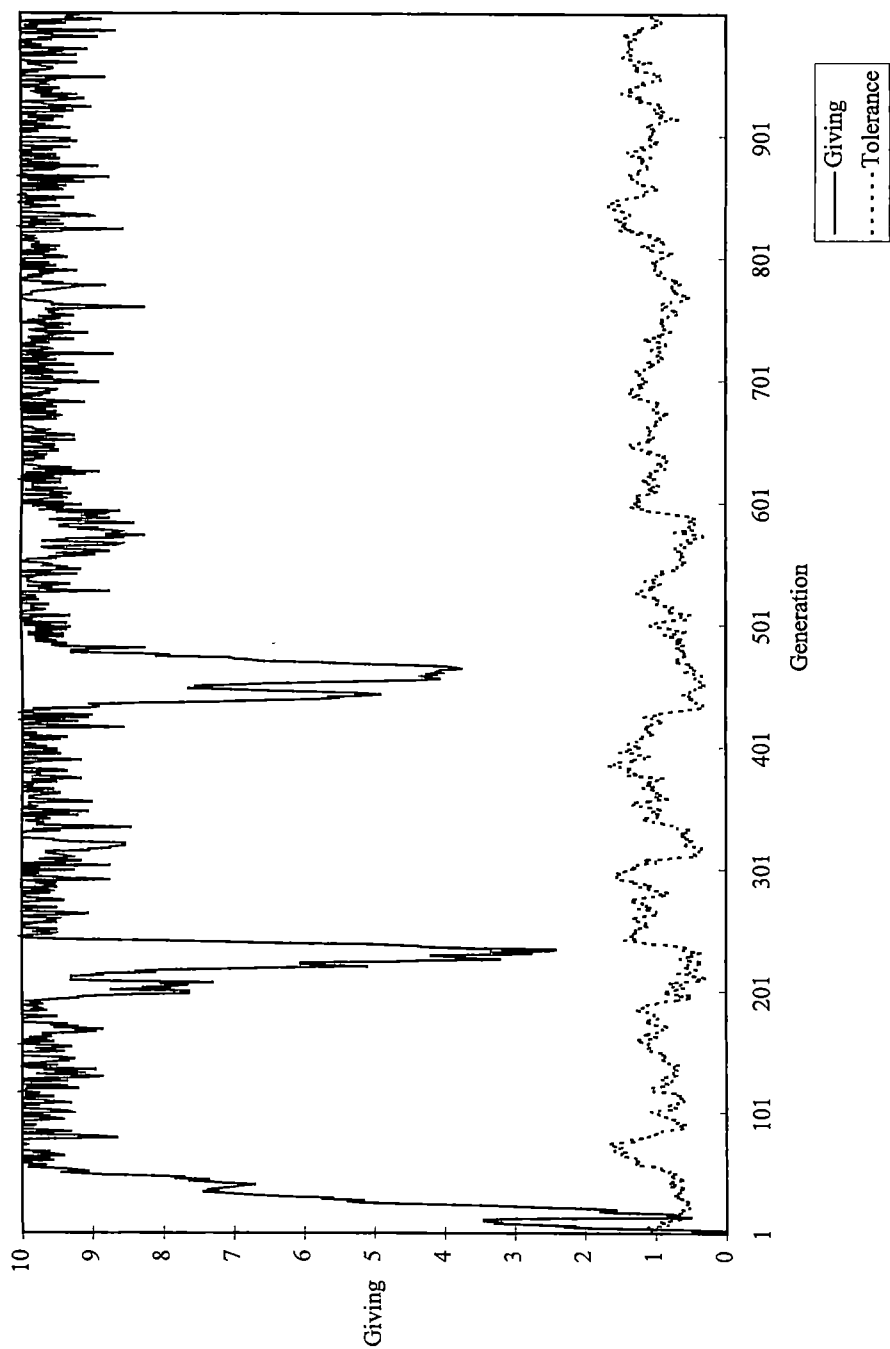


FIG. 3.—One example of simulation 1

not affect an actor's own total profit. Only the giving gene affects her profit. Therefore, this fluctuation should be random, and it was produced by mutation. Even with this fluctuation, pure-generalized exchange emerged. Therefore, the emergence of generalized exchange does not require a sense of fairness as a collective norm.

Simulation 2

The results of simulation 1 show that the emergence of generalized exchange is easier than had been expected. It does not require central authority, altruism, nor collective norms. However, the new model in this article requires another major assumption: perfect information. The new model can apply only in situations in which everybody knows everybody else. As long as this assumption exists, the emergence of generalized exchange remains possible only in limited situations. Therefore, in this second simulation, I loosen this assumption.

In natural settings, people are embedded in social networks, such as family, school, company, organization, and so forth. People interact with the other members of these networks more frequently than with those who are outside the networks (e.g., Grannis 1998; Marsden and Laumann 1984; Wellman and Berkowitz 1988). Therefore, a reasonable assumption is that each member of a society does not know all of the other members of a society but knows only other members of the networks that are significant to himself. In other words, perfect information may not hold true in an entire society but may hold true in these significant networks. However, even in this situation, it is anticipated that pure-generalized exchange can emerge. Based on simulation 1, I anticipated that generalized exchange can emerge first in a local small group. Because it is plausible that a large society consists of many small groups or networks, and they overlap to some extent, it is possible that generalized exchange then spreads beyond the specific group or network.

To test this prediction, I conducted a second series of simulations. The purpose of this simulation is to see whether generalized exchange can emerge even in a society in which each member knows only a part of the society. This simulation is also based on the evolutionary approach. To operationalize a large society that consists of many small local groups, I used the territorial system (Axelrod 1984). In this situation, each actor is embedded in a two-dimensional space and interacts only with his spatial neighbors. Although there are several other ways to represent a society in which each member knows only a part of it, I used the territorial system because it is the simplest way. According to Axelrod (1984), territories can be thought of in two ways. One way is in terms of geography and physical space. Each house or each country is located on the surface of the earth.

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

FIG. 4.—Territorial system. The number of each cell represents the number of each actor.

The other is in terms of an abstract space of characteristics. For example, people interact with relatively similar others more frequently than with relatively dissimilar others. The mechanism of spreading of successful strategies can occur in two ways, too. One is imitation, and the other is colonization. If the neighbor is doing well, the behavior of the neighbor can be imitated. Or, the location of a less successful actor can be taken over by an offspring of a more successful neighbor. But whether spreading occurs by imitation or colonization, the idea of a territorial system is the same: neighbors interact with each other, and the most successful actors spread to bordering locations.

In this simulation, a society consists of 100 actors. Each actor is embedded in a two-dimensional space. Figure 4 shows an example. Each actor knows only eight adjoining actors around herself.¹⁷ For example, actor 55

¹⁷ Figure 4 is considered to be a square. In other words, the number of actors that peripheral actors know is less than the number that central actors know. For example, actor 1 knows only actors 2, 11, and 12. Although the territorial system in Axelrod (1984) is a surface of a sphere, in order to guarantee that everyone has exactly the same number of neighbors, I decided to use a square for simplicity. Replicating this 50 times should be sufficient to cancel out the effect of peripheral actors.

TABLE 2
PARAMETERS OF SIMULATION 2

Parameter Names	Values
Group size	100
Number of replications	50
Number of generations	200
Number of trials per generation	10
Number of resources that each actor is given in each trial	10
Value of resources that each actor receives from others	2
Giving gene	from 0 to 10
Tolerance gene	from .1 to 2.0
Mutation rate01 (1%)

knows only actors 44, 45, 46, 54, 56, 64, 65, and 66. She does not know the other actors, such as actor 43. It means that each actor belongs to a small group of nine members from each actor's point of view. In this small local group, each actor does the same thing as in simulation 1. Table 2 shows the parameters. However, there is one difference between the algorithm of simulation 1 and that of simulation 2. In simulation 1, the criterion of "natural selection" is the mean and SD of total profit for the entire group. An actor whose total profit is large can have more "children" than low performers in each generation. However, in simulation 2, if one actor has any neighbor whose total profit was higher than his own profit, this actor learns the strategy of the most successful of them regardless of the mean or SD of the entire society. Figure 5 shows an illustrative example based on colonization. For simplicity, I assume no mutation occurred from generation t to generation $t + 1$ (but in the actual simulations, it occurs). For example, actor (9, 9), whose giving gene is 5 and whose total profit is 146 points, was taken over by actor (8, 9), whose giving gene is 7 and whose total profit is 334 points. This actor (8, 9) in generation t also colonized actor (9, 8) and actor (9, 10). However, this actor was taken over by actor (7, 9) whose total profit was 368 points.

In simulation 2, there is no independent variable. The initial value of the giving gene was set to 0. The initial value of the tolerance gene was determined randomly. The result is striking: The mean of the giving gene during the last 10 trials was 9.47, and SD was 1.77. Since the maximum potential value is 10.0, a mean of 9.47 implies that almost perfect generalized exchange emerged. There were only two replications in which the mean of the giving gene was under 9.5. Figure 6 shows an example of the history of the society in one replication. This figure shows that generalized

0	0	0	0	0	0	0	0	0	0
(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
0	0	0	0	0	0	0	0	0	0
(100)	(116)	(108)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
0	4	0	0	0	0	0	0	0	0
(100)	(60)	(108)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
0	0	0	0	0	0	0	0	0	0
(100)	(132)	(116)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
0	0	0	0	0	0	1	7	0	0
(100)	(100)	(100)	(100)	(100)	(100)	(90)	(68)	(100)	(100)
0	0	0	0	0	0	0	1	7	7
(100)	(100)	(100)	(100)	(100)	(100)	(100)	(102)	(258)	(100)
0	0	0	0	0	0	4	7	7	7
(100)	(100)	(100)	(100)	(100)	(100)	(100)	(236)	(368)	(240)
0	0	0	0	0	4	4	7	7	7
(100)	(100)	(100)	(100)	(100)	(68)	(116)	(320)	(334)	(146)
0	0	0	0	0	4	4	4	5	5
(100)	(112)	(118)	(106)	(100)	(148)	(108)	(68)	(146)	(130)
0	0	3	0	0	4	4	4	5	5
(100)	(118)	(70)	(106)	(100)	(76)	(116)	(68)	(124)	(80)

↓

0	0	0	0	0	0	0	0	0	0
(116)	(116)	(116)	(108)	(100)	(100)	(100)	(100)	(100)	(100)
0	0	0	0	0	0	0	0	0	0
(116)	(116)	(116)	(108)	(100)	(100)	(100)	(100)	(100)	(100)
0	0	0	0	0	0	0	0	0	0
(132)	(132)	(132)	(116)	(100)	(100)	(100)	(100)	(100)	(100)
0	0	0	0	0	0	0	0	0	0
(132)	(132)	(132)	(116)	(100)	(100)	(100)	(100)	(100)	(100)
0	0	0	0	0	0	1	7	7	7
(132)	(132)	(132)	(116)	(100)	(100)	(102)	(258)	(258)	(258)
0	0	0	0	0	0	7	7	7	7
(100)	(100)	(100)	(100)	(100)	(100)	(236)	(368)	(368)	(368)
0	0	0	0	0	4	7	7	7	7
(100)	(100)	(100)	(100)	(100)	(116)	(320)	(368)	(368)	(368)
0	0	0	0	4	4	7	7	7	7
(112)	(118)	(118)	(118)	(148)	(148)	(320)	(368)	(368)	(368)
0	0	0	0	4	4	7	7	7	7
(118)	(118)	(118)	(118)	(148)	(148)	(320)	(334)	(334)	(334)
0	0	0	0	4	4	4	5	5	5
(118)	(118)	(118)	(118)	(148)	(148)	(148)	(146)	(146)	(146)

FIG. 5.—An example of colonization. The top table shows the situation at generation t , and the bottom table shows the situation at generation $t + 1$. The numbers outside the parentheses represent the value of the giving gene. The number inside the parentheses represent the total profit.

exchange is very stable and never declines. This implies that the situation in which everyone has a high level of the giving gene is an equilibrium. Again, the mean of the tolerance gene fluctuates as expected. There seems to be no equilibrium point. From these results, we can conclude that pure-generalized exchange can emerge even in a society in which each member has information only about her neighbors. Therefore, the assumption of perfect information in the new model is successfully loosened.

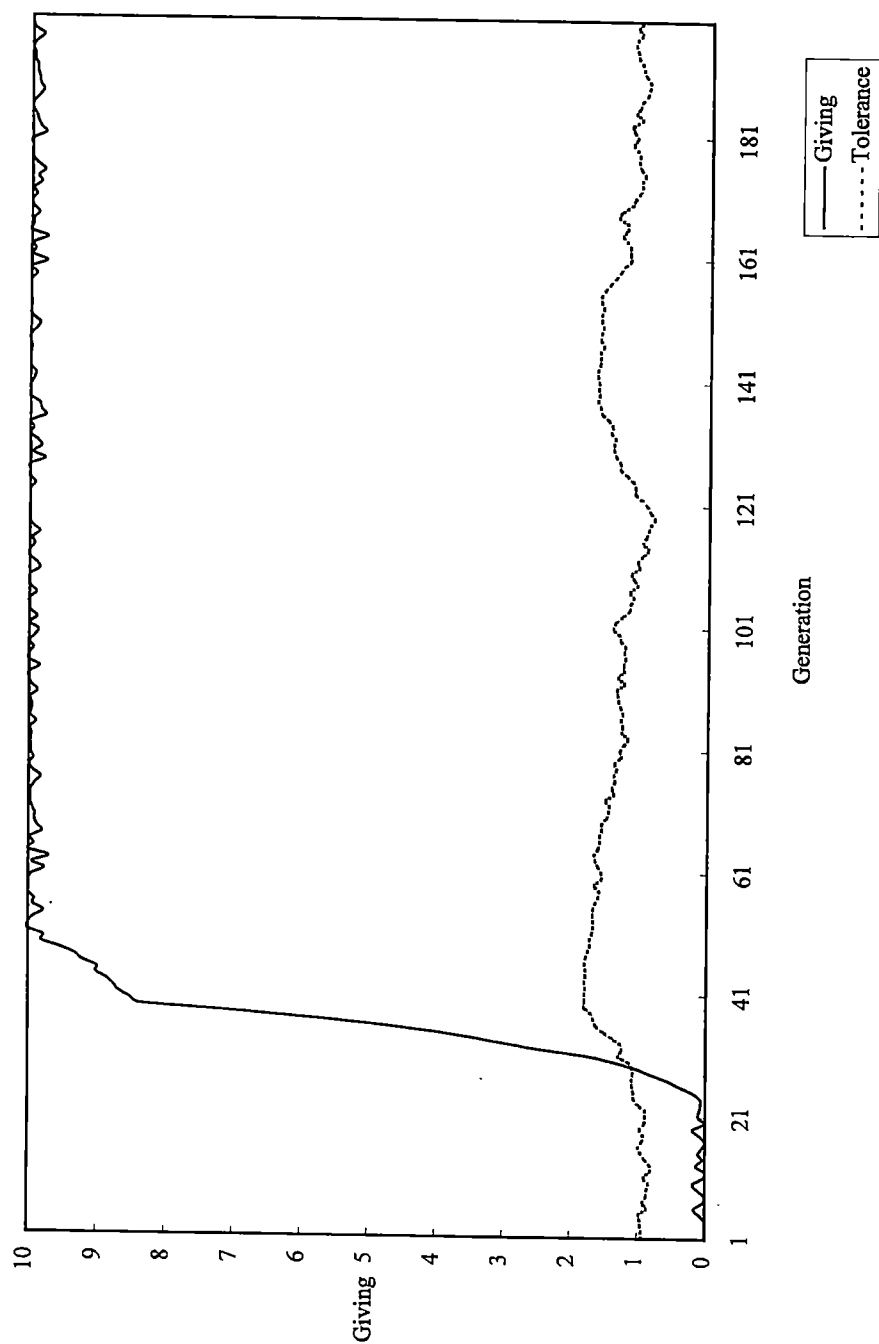


FIG. 6.—One example of simulation 2

DISCUSSION

Generalized exchange has been explained by altruism (e.g., Sahlins 1972; Nye 1979) or collective norms (e.g., Ekeh 1974; Lévi-Strauss 1949). However, if we assume that actors are self-interested, the emergence of generalized exchange, characterized by unilateral resource giving, has posed a puzzle because of the possibility of free riding. Thus, we need to solve the social dilemma problem that is involved in a generalized exchange situation. One of the main solutions for the free rider problem is selective incentives (Olson 1965), but that introduces another issue: the second-order free rider problem (Oliver 1980; Yamagishi 1986*a*, 1986*b*). Previous studies that adopt the social exchange/rational choice perspective have solved these problems by imposing a particular social structure (i.e., network-generalized exchange; e.g., Boyd and Richerson 1989; Yamagishi and Cook 1993). However, this solution has its own weakness and limitation. This study also adopted the social exchange/rational choice approach and tried to impose particular social structures as little as possible.

In order to pursue this goal, this study proposed a new model that can solve the free rider problem that exists in generalized exchange. I first identified a new situation, pure-generalized exchange, in which each actor gives resources to the recipient(s) of her choice. Second, I proposed a new strategy, the fairness-based selective-giving strategy, in which actors select recipients whose behaviors satisfy their own criteria of fairness. The main argument was that this fairness-based selective-giving strategy would make pure-generalized exchange possible.

To show that this normative and theoretical argument can hold, I conducted two evolutionary simulations. The results of simulation 1 showed that pure-generalized exchange can emerge very easily, even in a society in which no one gives to anyone else initially (i.e., a society of egoists). They also showed that pure-generalized exchange can emerge even in a society in which members have different standards of fairness. Therefore, altruism and a collective sense of fairness are no longer required. However, in simulation 1, this solution required another strong assumption, the assumption of perfect information. In order to select the recipients, each actor must know every other actor's behavior.

To loosen this assumption as much as possible, I conducted simulation 2. Using the territorial system, simulation 2 created societies where members had information only about their immediate neighbors. The results again supported the theoretical argument. Even in this situation, pure-generalized exchange could emerge by virtue of the fairness-based selective-giving strategy.

In sum, the results of these simulations show that we can explain the emergence of pure-generalized exchange even if we do not assume pre-

existing altruistic motivation or norm of reciprocity, provided that each individual has a sense of fairness.¹⁸

Because I assume self-interested actors, it might be possible to interpret the results of the simulations as showing that actors intentionally give their resources unilaterally in order to increase their own profits. This interpretation corresponds to the forward-looking rationality that typical rational choice theory adopts. However, there is another reasonable interpretation. Based on the logic of learning theory and evolutionary biology, it is possible that actors come to give their resources unilaterally because unilateral resource giving brings profit (but not necessarily from the recipient)—regardless of the actors' intentions. This corresponds to backward-looking rationality. Because pure-generalized exchange is complex, and because there is no guarantee of *quid pro quo*, actors may not know the consequences of their behavior in advance. Thus, in a society in which pure-generalized exchange holds, actors who give their resources altruistically (i.e., not expecting future return) to others receive benefits. Therefore, the new model in this article may suggest a foundation for "altruism" (i.e., certain behavioral patterns that encourage people to believe that an actor might be an altruistic person, but not altruistic motivation *per se*).¹⁹ In English, there are several interesting proverbs, such as "Charity is a good investment," or "He who gives to another bestows on himself." Similar proverbs exist across the world. They might represent the truth. The computer simulation, of course, is silent on the question of which motivation (i.e., maximizing one's outcome or altruism) is empirically true. What I have shown is that behaviors that look like altruism can emerge even if we do not assume altruism in the first place.

However, this "altruism" is not altruism in a broad sense. Indeed, if this study could show a foundation for altruism, it would not be a foundation for "universal altruism," but a foundation for "selective altruism" or "discriminating altruism" (Hardin 1982; Takagi 1994, 1996). Universal altruism means no discrimination. A person adopting universal altruism gives resources to anybody, whether the recipient is a good person or a

¹⁸ We can consider an alternative model that uses the principle of homophily; people give to those who are similar to themselves in level of giving, not to those who give less or to those who give more. This may work as well as fairness-based selective-giving strategy. This is certainly an interesting approach for the future research. One could also specify the conditions under which generalized exchange emerges. This is beyond the scope of this article; however, I can speculate several effects of the conditions on generalized exchange. I believe that the density of the network in simulation 2 should have a positive effect on cooperation. The more eyes there are, the more likely cooperation is to be achieved. And, I believe that the size of groups does not matter in simulation 1.

¹⁹ Takagi (1994, 1996) also suggested this implication.

killer. In contrast, selective altruism means that a person behaves altruistically to one kind of person but not to another kind. A person who adopts discriminating altruism selects a recipient based on certain criterion. The extreme case is kin selection in evolutionary biology. In this case, the selection criterion is the number of common genes between the person and the target. What Hardin (1982) originally had in mind is altruism only to in-group members (e.g., cronyism, tribalism, and patriotism). In this case, the criterion might be race, ethnicity, or country. Although the context is different, the concept of the fairness-based selective-giving strategy fits that of discriminating altruism. In this case, the criterion is how a target behaves to others compared to how oneself behaves to others. I called this fairness in this article. Therefore, what this study may show is a foundation of selective altruism based on a type of fairness. This is parallel to what Axelrod (1984) showed. He showed that reciprocal altruism between dyads can emerge even if we assume self-interested actors. This article showed that selective altruism based on fairness among multiple actors can emerge even if we assume self-interested actors.

Limits and Future Directions

The new model that I proposed in this article can explain generalized exchange under much less restrictive conditions than those used in previous research. However, it cannot explain all patterns of generalized exchange. First, the fairness-based selective-giving strategy requires a certain amount of information about potential recipients. Based on this information, an actor using fairness-based selective giving decides whether or not to give. Although this condition was loosened in simulation 2, at the same time, the potential recipients of each actor were only his neighbors. Therefore, this model cannot explain unilateral resource giving to a complete stranger (e.g., helping a stranded driver), although this model is appropriate for Ekeh's (1974) example of witnessing a burglar.

From this restriction, two future directions can be considered. First, it might be possible that generalized exchange emerges originally in a relatively small subgroup (i.e., neighbors, friends, work-related acquaintances, etc.) and then spreads throughout an entire society. The fact that cultural anthropologists have observed generalized exchange in primitive societies might suggest this possibility. In order to examine this possibility, either social network research (e.g., Granovetter 1973; Blau 1977; Marsden 1987), which examines the internal structure of a society, or evolutionary biology (e.g., Hamilton 1964), which examines altruism in kinship groups, would be useful.

The second direction involves studies about behavior under uncertainty. Following the new model in this article, if an actor gives her re-

sources to a stranger, she should estimate whether this stranger typically gives much or little. Very recently, research studying the relationship between social uncertainty and cognitive traits has begun. For example, Frank (1988) and Frank, Gilovich, and Regan (1993) argue that people can detect their partners' intentions to some extent in a PD game. Also, this line of research has begun to use the approach of the signaling game from evolutionary biology (e.g., Macy and Skvoretz 1998). Orbell and Dawes (1991, 1993) argue for a "false consensus" effect, that is, a person who is cooperative estimates that others also would be cooperative. Recent development of the theories of trust by Yamagishi and his associates (e.g., Yamagishi and Yamagishi 1994; Yamagishi, Cook and Watabe 1998) argue that people use their own general trust as a default to estimate others' benevolence. Linking generalized exchange to these findings would be very fruitful.

The other restriction of this study is the assumption of an individual sense of fairness. As I discussed above, the conception of fairness in this article is somewhat different from that used by previous fairness/justice researchers. Previous researchers have typically assumed a situation where fairness-based action directly affects one's own outcome. Certainly, people have an egocentric bias of the fairness judgment (e.g., Messe, Hymes, and MacCoun 1986; Messick and Sentis 1979). However, the sense of fairness in this article *does not* affect one's outcome. Having this sense of fairness does not fit either the perspective of forward-looking rationality that is typically assumed in rational choice and game theory or perspective of the backward-looking rationality that is typically assumed in social exchange theory, learning theory, and evolutionary theory (Heath 1976; Macy 1993a). Having a sense of fairness is not beneficial to the actor. As I argued above, it is true that the fairness-based selective-giving strategy can fit both perspectives, but the sense of fairness itself does not. However, having this sense of fairness does not contradict the rationality principles, either, because having it does not give any disadvantage to the actors. More technically, each person's profit is determined by the giving gene, not by the tolerance gene. As long as persons give to someone else, they are likely to be targets of giving by others, whether they are selective or not. That is why the mean of the tolerance gene fluctuated around 1.00. Therefore, this study does not show a foundation for fairness, although it would have been better if it could. Thus, I leave this as an assumption: a certain level of fairness within a society is a necessary condition for the emergence of generalized exchange, although the fairness criterion of each individual can be different. At this point, I can only suggest that there might be an interdependency between a tendency to give and fairness. One preliminary experimental study of generalized exchange showed that there is a positive correlation between the amount of giving and the con-

cern with fairness (Takahashi and Yamagishi 1999). In other words, empirically, generous givers are more selective (choosy) than stingy givers. Clearly future research on the origin of the sense of fairness is needed.

REFERENCES

- Adams, J. Stacy. 1965. "Inequity in Social Exchange." *Advances in Experimental Social Psychology* 2:267-99.
- Axelrod, R. 1980a. "Effective Choice in the Prisoner's Dilemma." *Journal of Conflict Resolution* 24:3-25.
- . 1980b. "More Effective Choice in the Prisoner's Dilemma." *Journal of Conflict Resolution* 24:379-403.
- . 1984. *The Evolution of Cooperation*. New York: Basic Books.
- . 1986. "An Evolutionary Approach to Norms." *American Political Science Review* 80:1095-111.
- Batson, C. D. 1987. "Prosocial Motivations: Is It Ever Truly Altruistic?" *Advances in Experimental Social Psychology* 20:65-122.
- Bearman, P. 1997. "Generalized Exchange." *American Journal of Sociology* 102:1383-415.
- Befu, H. 1977. "Social Exchange." *Annual Review of Anthropology* 6:255-81.
- . 1980. "Structural and Motivational Approaches to Social Exchange." Pp. 197-214 in *Social Exchange: Advances in Theory and Research*, edited by K. J. Gergen, M. S. Greensberg, and R. H. Willis. New York: Plenum Press.
- Bienenstock, Elisa Jayne, and Phillip Bonacich. 1992. "The Core as a Solution to Exclusionary Networks." *Social Networks* 14:231-343.
- Blau, P. M. 1964. *Exchange and Power in Social Life*. New York: Wiley.
- . 1977. *Inequality and Heterogeneity*. New York: Free Press.
- Boyd, R., and P. J. Richerson. 1989. "The Evolution of Indirect Reciprocity." *Social Networks* 11:213-36.
- Cook, Karen S., and Richard M. Emerson. 1978. "Power, Equity and Commitment in Exchange Networks." *American Sociological Review* 43:721-39.
- Cook, Karen S., Richard M. Emerson, Mary R. Gillmore, and Toshio Yamagishi. 1983. "The Distribution of Power in Exchange Networks: Theory and Experimental Results." *American Journal of Sociology* 89:275-305.
- Dawes, Robyn M. 1980. "Social Dilemmas." *Annual Review of Psychology* 31:169-93.
- Ekeh, P. P. 1974. *Social Exchange Theory: The Two Traditions*. Cambridge, Mass.: Harvard University Press.
- Emerson, R. M. 1972a. "Exchange Theory, Part I." Pp. 38-57 in *Sociological Theories in Progress*, vol. 2. Edited by J. Berger, M. Zelditch, and B. Anderson. Boston: Houghton-Mifflin.
- . 1972b. "Exchange Theory, Part II." Pp. 58-87 in *Sociological Theories in Progress*, vol. 2. Edited by J. Berger, M. Zelditch, and B. Anderson. Boston: Houghton-Mifflin.
- . 1976. "Social Exchange Theory." *Annual Review of Sociology* 2:335-62.
- . 1981. "Social Exchange Theory." Pp. 30-65 in *Social Psychology: Sociological Perspectives*, edited by M. Rosenberg and R. H. Turner. New York: Basic.
- Frank, R. H. 1988. *Passions within Reason*. New York: Norton.
- Frank, R. H., T. Gilovich, and D. T. Regan. 1993. "The Evolution of One-Shot Cooperation: An Experiment." *Ethology and Sociobiology* 14:247-56.
- Friedkin, Noah E. 1992. "An Expected Value Model of Social Power: Predictions for Selected Exchange Networks." *Social Networks* 14:213-29.

- Gillmore, Mary R. 1987. "Implications of Generalized versus Restricted Exchange." Pp. 170-89 in *Social Exchange Theory*, edited by Karen S. Cook. Newbury Park, Calif.: Sage.
- Grannis, Rick. 1998. "Trivial Streets and Residential Segregation." *American Journal of Sociology* 103:1530-64.
- Granovetter, M. S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78:1360-80.
- Hamilton, W. D. 1964. "The Genetic Theory of Social Behavior I and II." *Journal of Theoretical Biology* 7:1-52.
- Hardin, Garret. 1982. "Discriminating Altruisms." *Zygon* 17:163-86.
- Hayashi, N. 1995. "Emergence of Cooperation in One-Shot Prisoner's Dilemmas and the Role of Trust." *Japanese Journal of Psychology* 66:184-90 (in Japanese).
- Hayashi, N., N. Jin, and T. Yamagishi. 1993. "Prisoner's Dilemma Network: A Computer Simulation of Strategies." *Research in Social Psychology* 8:33-43 (in Japanese).
- Hayashi, N., and T. Yamagishi. 1998. "Selective Play: Choosing Partners in an Uncertain World." *Personality and Social Psychology Review* 2:276-89.
- Heath, A. 1976. *Rational Choice and Social Exchange: A Critique of Exchange Theory*. Cambridge: Cambridge University Press.
- Heckathorn, Douglas D. 1988. "Collective Sanctions and the Creation of Prisoner's Dilemma Norms." *American Journal of Sociology* 94:535-62.
- . 1989. "Collective Action and the Second Order Free Rider Problem." *Rationality and Society* 1:78-100.
- . 1990. "Collective Sanctions and Compliance Norms: A Formal Theory of Group-Mediated Social Control." *American Sociological Review* 55:366-84.
- . 1993. "Collective Action and Group Heterogeneity: Voluntary Provision versus Selective Incentives." *American Sociological Review* 58:329-50.
- . 1996. "The Dynamics and Dilemmas of Collective Action." *American Sociological Review* 61:250-77.
- Homans, G. C. (1961) 1974. *Social Behavior: Its Elementary Forms*. New York: Harcourt Brace.
- Jasso, G. 1980. "A New Theory of Distributive Justice." *American Sociological Review* 45:3-32.
- Lawler, E. J. 1986. "Bilateral Deterrence and Conflict Spiral: A Theoretical Analysis." Pp. 107-30 in *Advances in Group Processes*, vol. 3. Edited by Edward J. Lawler. Greenwich, Conn.: JAI Press.
- Lévi-Strauss, C. 1949. *Les structures élémentaires de la parenté*. Paris: Presses Universitaires de France.
- Macy, M. 1990. "Learning Theory and the Logic of Critical Mass." *American Sociological Review* 55:809-26.
- . 1991a. "Chains of Cooperation: Threshold Effects in Collective Action." *American Sociological Review* 56:730-47.
- . 1991b. "Learning to Cooperate: Stochastic and Tacit Collusion in Social Exchange." *American Journal of Sociology* 97:808-43.
- . 1993a. "Backward-Looking Social Control." *American Sociological Review* 58:819-36.
- . 1993b. "Social Learning and the Structure of Collective Action." *Advances in Group Processes* 10:1-36.
- Macy, Michael W., and John Skvoretz. 1998. "The Evolution of Trust and Cooperation between Strangers: A Computational Model." *American Sociological Review* 63:638-60.
- Malinowski, B. 1922. *Argonauts of the Western Pacific*. London: Routledge & Kegan Paul.
- Markovsky, B. 1985. "Toward a Multilevel Distributive Justice Theory." *American Sociological Review* 50:822-39.

- Markovsky, B., D. Willer, and T. Patton. 1988. "Power Relations in Exchange Networks." *American Sociological Review* 53:220-36.
- Marsden, P. V. 1987. "Core Discussion Networks of Americans." *American Sociological Review* 52:122-33.
- Marsden, P. V., and E. Laumann. 1984. "Mathematical Ideas in Social Structural Analysis." *Journal of Mathematical Sociology* 10:271-94.
- Mauss, Marcel. (1925) 1954. *The Gift: Forms and Functions of Exchange in Archaic Societies*. London: Cohen & West Ltd.
- Messe, Lawrence A., Robert W. Hymes, and Robert J. MacCoun. 1986. "Group Categorization and Distributive Justice Decisions." Pp. 227-48 in *Justice in Social Relations*, edited by H. Bierhoff, R. L. Cohen, and J. Greenberg. New York: Plenum.
- Messick, David M., and Marilynn B. Brewer. 1983. "Solving Social Dilemmas: A Review." Pp. 11-14 in *Review of Personality and Social Psychology*, vol. 4. Edited by L. Wheeler. Beverly Hills, Calif.: Sage.
- Messick, D. M., and K. P. Sentis. 1979. "Fairness and Preference." *Journal of Experimental Social Psychology* 15:416-34.
- Molm, Linda D. 1994. "Dependence and Risk: Transforming the Structure of Social Exchange." *Social Psychology Quarterly* 57:163-76.
- Molm, L. D., and K. S. Cook. 1995. "Social Exchange and Exchange Networks." Pp. 209-35 in *Sociological Perspectives on Social Psychology*, edited by K. S. Cook, G. A. Fine, and J. S. House. Boston: Allyn & Bacon.
- Nye, F. Ivan. 1979. "Choice, Exchange and the Family." Pp. 1-41 in *Contemporary Theories about the Family*, vol. 2. Edited by Wesley R. Burr, Reuben Hill, F. Ivan Nye, and Ira L. Reiss. New York: Free Press.
- Oliver, Pamela. 1980. "Reward and Punishments as Selective Incentives for Collective Action: Theoretical Investigations." *American Journal of Sociology* 85:1356-75.
- Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, Mass.: Harvard University Press.
- Orbell, John, and Robyn M. Dawes. 1981. "Social Dilemmas." Pp. 37-65 in *Progress in Applied Social Psychology*, vol. 1. Edited by G. Stephenson and J. H. Davis. Chichester: Wiley.
- . 1991. "A 'Cognitive Miser' Theory of Cooperators' Advantage." *American Political Science Review* 85:515-28.
- . 1993. "Social Welfare, Cooperators' Advantage and the Option of not Playing the Game." *American Sociological Review* 58:878-800.
- Oskamp, S. 1971. "Effects of Programmed Strategies on Cooperation in the Prisoner's Dilemma and Other Mixed-Motive Strategies." *Journal of Conflict Resolution* 15: 225-69.
- Piliavin, J. A., and H. W. Charng. 1990. "Altruism: A Review of Recent Theory and Research." *Annual Review of Sociology* 16:27-65.
- Sahlins, Marshall. 1972. *Stone Age Economics*. Chicago: Aldine.
- Schuessler, R. 1989. "Exit Threats and Cooperation under Anonymity." *Journal of Conflict Resolution* 33:728-49.
- Sesardic, Neven. 1995. "Recent Work on Human Altruism and Evolution." *Ethics* 106:128-57.
- Stroebe, Wolfgang, and Bruno S. Frey. 1982. "Self-Interest and Collective Action: The Economics and Psychology of Public Goods." *British Journal of Social Psychology* 21:121-37.
- Takagi, E. 1994. "The Evolutionary Giving Game: A Computer Simulation Paradigm for the Study of Social Exchange." *Annals of Saitama University* 30:23-55 (in Japanese).
- . 1996. "The Generalized Exchange Perspective on the Evolution of Altruism." Pp. 311-36 in *Frontiers in Social Dilemmas Research*, edited by W. B. G. Liebrand and D. M. Messick. Berlin: Springer-Verlag.

- Takahashi, N., and T. Yamagishi. 1995. "Altruism Can Be Profitable if People Are Not Too Generous or Too Demanding." Paper presented at the 6th International Conference on Social Dilemmas, NIAS, Wassenaar, The Netherlands, June 18–23.
- . 1996. "Social Relational Foundations of Altruistic Behavior." *Japanese Journal of Experimental Social Psychology* 36:1–11 (in Japanese).
- . 1999. "Voluntary Formation of a Generalized Exchange System: An Experimental Study of Discriminating Altruism." *Japanese Journal of Psychology* 70:9–16 (in Japanese).
- Thibaut, J. W., and H. H. Kelly. 1959. *The Social Psychology of Groups*. New York: Wiley.
- Uehara, Edwina. 1990. "Dual Exchange Theory, Social Networks, and Informal Social Support." *American Journal of Sociology* 96:521–57.
- Vanberg, V., and R. Congleton. 1992. "Rationality, Morality and Exit." *American Political Science Review* 86:418–31.
- Walster, Elaine G., William Walster, and Ellen Berscheid. 1978. *Equity: Theory and Research*. Allyn & Bacon.
- Wellman, B., and S. D. Berkowitz. 1988. *Social Structures: A Network Approach*. New York: Cambridge University Press.
- Wilson, W. 1971. "Reciprocation and Other Techniques for Inducing Cooperation in the Prisoner's Dilemma Game." *Journal of Conflict Resolution* 15:167–95.
- Yamagishi, Toshio. 1986a. "The Provision of a Sanctioning System as a Public Good." *Journal of Personality and Social Psychology* 51:110–16.
- . 1986b. "The Structural Goal/Expectation Theory of Cooperation in Social Dilemmas." Pp. 51–87 in *Advances in Group Processes*, vol. 3. Edited by E. J. Lawler. Greenwich, Conn.: JAI Press.
- . 1989a. "Major Theoretical Approaches in Social Dilemmas Research." *Japanese Psychological Review* 33:64–96 (in Japanese).
- . 1989b. "Unintended Consequences of Some Solutions to the Social Dilemmas Problem." *Sociological Theory and Method* 4:21–47 (in Japanese).
- . 1995. "Social Dilemmas." Pp. 209–35 in *Sociological Perspectives on Social Psychology*, edited by K. S. Cook, G. A. Fine, and J. S. House. Boston: Allyn & Bacon.
- Yamagishi, T., and K. S. Cook. 1993. "Generalized Exchange and Social Dilemmas." *Social Psychology Quarterly* 56:235–48.
- Yamagishi, T., K. S. Cook, and M. Watabe. 1998. "Uncertainty, Trust, and Commitment Formation in the United States and Japan." *American Journal of Sociology* 104:165–94.
- Yamagishi, Toshio, Mary R. Gillmore, and Karen S. Cook. 1988. "Network Connections and the Distribution of Power in Exchange Networks." *American Journal of Sociology* 93:833–51.
- Yamagishi, T., and N. Hayashi. 1996. "Selective Play: Social Embeddedness of Social Dilemmas." Pp. 363–84 in *Frontiers in Social Dilemmas Research*, edited by W. B. G. Liebrand and D. M. Messick. Berlin: Springer-Verlag.
- Yamagishi, T., N. Hayashi, and N. Jin. 1994. "Prisoner's Dilemma Network: Selection Strategy versus Action Strategy." Pp. 233–50 in *Social Dilemmas and Cooperation*, edited by U. Schulz, W. Albers, and U. Mueller. Berlin: Springer-Verlag.
- Yamagishi, T., and M. Yamagishi. 1994. "Trust and Commitment in the United States and Japan." *Motivation and Emotion* 18:129–66.
- Ziegler, R. 1990. "The *Kula*: Social Order, Barter, and Ceremonial Exchange." Pp. 141–68 in *Social Institutions: Their emergence, Maintenance and Effects*, edited by M. Hechter, K-D Opp, and R. Wippler. New York: Aldine de Gruyter.

Economic Transformation and Income Inequality in Urban China: Evidence from Panel Data¹

Xueguang Zhou
Duke University

Using panel data of 4,730 urban residents drawn from 20 cities in China, this article examines changes in income determinants between the prereform and reform eras. To guide this empirical study, a conceptual model is developed that emphasizes the coevolution of politics and markets to synthesize theoretical ideas in the recent debate on the transformation of state socialist societies. The findings show significant changes in returns to education and in the rise of private/hybrid firms in the reform era. There is also strong evidence of institutional persistence in returns to positional power and in the organizational hierarchy. These findings reveal multifaceted processes of transformation that call for more sophisticated theoretical models and in-depth institutional analyses.

INTRODUCTION

An ongoing debate in recent studies of the transformation of state socialist societies centers on how to explain (and predict) the processes and mechanisms of large-scale social changes and their implications for the social stratification order in these societies (see Symposium on Market Transition in *American Journal of Sociology*, vol. 101, no. 4). On the one hand, Victor Nee's market transition theory (1989, 1991, 1996) emphasizes the role of emerging markets that compete with and undermine state socialist redistributive economy and alter the stratification order. On the other hand, other scholars emphasize various aspects of the political economy

¹ This research is supported by a grant from the National Science Foundation (SBR-9413540), an ASA/NSF fund for the advancement of the discipline, and a Spencer Fellowship. I am grateful to the *AJS* reviewers for their comments on earlier versions of this article. I thank Liren Hou for his able research assistance and the departments of sociology at Fudan University and the People's University, the Institute of Sociology at Tianjin Academy of Social Sciences, and in particular, Weida Fan, Qiang Li, Yunkang Pan, and Xizhe Peng for their assistance in data collection. Direct correspondence to Xueguang Zhou, Department of Sociology, Duke University, Durham, North Carolina 27708. E-mail: xzhou@soc.duke.edu

of transformation in which existing political and economic institutions mold emerging economic institutions.

The main issues under debate are closely related to long-standing sociological interests in explaining social changes and processes of social stratification. They also reflect sociologists' renewed interests in the role of social and political institutions in economic activities and how social institutions persist and change. It is not surprising that comparative sociologists have embraced the opportunity and that a large number of studies appeared in major sociological journals addressing these issues (Bian and Logan 1996; Gerber and Hout 1998; Guthrie 1997; Peng 1992; Róna-Tas 1994; Stark 1996; Walder 1995*a*, 1995*b*; Xie and Hannum 1996; Zhou, Tuma, and Moen 1997).

But empirical evidence is far from consistent and conclusive in adjudicating between competing theoretical hypotheses. The state of the field is partly dictated by the complexities and transitional characteristics of the transformation processes that evolve over time. From a researcher's point of view, however, these unsettling issues also reflect weaknesses in research design, data quality, and conceptual framework.

The accumulation of knowledge particularly suffers from a lack of consensus on the operationalization of key concepts. For instance, the "cadre" status has been used as an indicator of the key concept of "positional power" in state socialist redistribution. But it was measured differently in different studies, even in similar urban or rural settings. As Parish and Michelson (1996) showed, different operationalizations of the "cadre" status may lead to very different conclusions about the role of positional power, even when analyzing the same data set.

Although assessment of changes over time has been the main focus of research, almost all studies have used cross-sectional research designs. Such a design facilitates particular research purposes in comparing interregion or intergroup differences, but they are inadequate for studying changes *over time*. In addition, a cross-sectional research design is ill equipped to deal with unobserved individual attributes that may confound several key theoretical issues. For instance, the latent attributes of individual ability may affect an individual's educational level and occupational status, as well as his or her party membership.

Substantively, the debate in part stems from different emphases given to different processes of social changes. Like any large-scale social changes, the transformation of state socialism involves multifaceted processes, in which the expansion of markets is one, albeit a very important one, of many. Such transformations inevitably produce a variety of patterns of changes that defies a single theoretical logic. In this light, there is room for a theoretical synthesis (Szelényi and Kostello 1996).

In this study, I take a step to address these issues. I propose a conceptual

model that emphasizes the interplay between politics and markets as a coevolutionary process and advocates a focus on the *processes* of institutional change that provides a basis for a theoretical synthesis among competing arguments. My analytical focus is on a confirmatory study based on the multiple processes identified in the literature, reinterpreted in light of the proposed model. I report an analysis of changes in income determinants based on a panel data of 4,730 urban residents drawn from 20 Chinese cities, whose income information was collected retrospectively for selected years between 1955 and 1994. My focus on urban China, where China's redistributive economy has reached its most elaborate form, provides a more appropriate examination of the changes in state socialist redistribution. The use of panel data allows one to assess changes over time in a more satisfactory way and helps alleviate some potential problems in measurements of key concepts.

A caveat is in order. Income is but one indicator of economic benefits associated with positions. Recently, scholars (Oberschall 1996; Walder 1992; Zhou et al. 1997) called attention to the importance of latent and nonmonetary resources distributed through workplaces and other channels. I focus on income determinants for two reasons: First, income distribution is sensitive to changes in economic institutions, and a study of income determinants can detect the most salient changes in the transformation processes. Second, most empirical studies thus far have examined patterns of income inequality. My study addresses the same set of empirical questions and can be directly compared with previous studies in the literature.

EXPLAINING INSTITUTIONAL CHANGES IN CHINA: THEORETICAL ISSUES

The Theoretical Debate

Based on a series of studies of income inequality in rural China, Nee's market transition theory (1989, 1991, 1992, 1996) was one of the earliest attempts to theorize about sources of social changes in transitions from state socialism and their implications for changes in the social stratification order. Nee's bold theoretical statements and insistent defense of his positions have stimulated active research and debates on both alternative theoretical explanations and empirical assessments of the competing theoretical arguments.

Central to market transition theory is its emphasis on the importance of emerging market economies. In Nee's view, markets and state socialist redistribution represent two fundamentally different logics of resource allocation. The emergence of a market economy introduces new mechanisms of resource allocation that challenge and undermine state socialist

redistributive institutions. Nee's earlier arguments (1989, 1991) predicted two fundamental changes: First, the emergence of a market economy alters the opportunity structure shaped by a redistributive economy. Second, changes in mechanisms of resource allocation lead to changes in social stratification order.

On the other hand, other scholars have developed several alternative, competing explanations. These explanations consist of several streams of arguments emphasizing the increasing role of local governments (Walder 1995*b*), the persistence of political power (Bian and Logan 1996), the conversion of political power to economic resources (Róna-Tas 1994), changes in "political markets" (Parish and Michelson 1996), institutional reconfiguration (Stark 1996), and the institutional arrangements of work organizations (Zhou et al. 1997). Although they differ in emphases, these arguments share the main disposition that the ongoing transformation processes and emerging economic institutions are circumscribed by and intertwined with existing political, economic, and social institutions. Thus, an explanation of the sources and directions of the transformation must take seriously existing institutional arrangements and explain how they interact with and shape emerging new institutions.

A central issue in the debate is the role of institutions in theories of social changes. Nee's focus is on new market institutions whose advance forges new interests and relentlessly pushes aside, if not dismantles, "old" economic and political institutions. On this account, he criticized his critics as to "emphasize structural continuity with unreformed state socialism" and to focus "only on changing incentives for political actors in the state organizational hierarchy" (Nee 1996, pp. 914–15). Yet a similar critique can be made of market transition theory. Despite the claim to "incorporate the state as a potent causal force" (p. 916), Nee's work has made little effort to consider how the *interactions* between existing political and economic institutions and emerging institutions affect patterns of social stratification. For instance, Nee (1996) advocated an institutional analysis of market transition in China, but the path of change he emphasized is a process in which "power—control over resources—shifts progressively from political disposition to market institutions" (p. 910). Nee's arguments leave the impression that the entire outcomes of *institutional changes* can be explained by the advance of market economies, despite the resistance of "old" institutions.

The second major difference is related to the role of agency and interests. Nee emphasized the importance of new interests associated with emerging markets. As Nee (1996) argued: "Whereas opportunities for advancement were previously centered solely on decisions made by the redistributive bureaucracy and within the economy controlled by it, markets open up alternative avenues for mobility through emergent entrepreneur-

ship and labor markets." Along with these changes, "economic actors strive to institute new rules of competition and cooperation that serve their interests, both through informal arrangements and through formal institutional channels" (p. 910).

In contrast, the emphases on existing social and political institutions have led other scholars to recognize the importance of incumbents and vested interests in the transformation processes. The prevalence of political authority, bureaucratic bargaining, and governmental intervention in risk and benefit sharing has been noted in several studies and commentaries (Bian and Logan 1996; Naughton 1995; Oberschall 1996; Oi 1992; Parish and Michelson 1996; Shirk 1993; Walder 1995*b*, 1996). Thus, the incumbents may use their positional power and privileged access to decision-making processes to influence state policies and governmental regulations in favor of their interests. Or they may be in an advantageous position to capitalize on opportunities and economic benefits, relative to other social groups. Therefore, their positional power and advantage may persist even when the role of market institutions increases.²

Coevolution between Politics and Markets: Sketch of a Conceptual Model

As a basis for a theoretical synthesis, I propose a conceptual model that treats institutional changes as processes of coevolution between politics and markets. By "politics," I refer to patterns of interest articulation in the political arena associated with both current and emerging political and economic institutions. By "markets," I refer to the modes of resource allocation and economic transactions that take place through price systems and involve "autonomous" economic agents. The interplay between politics and markets has been emphasized by several scholars (see Bian and Logan 1996; Guthrie 1997; Parish and Michelson 1996; Róna-Tas 1994; Stark 1996; Walder 1996). I highlight some theoretical considerations along this line of reasoning.

Using the proposed model, I intend to portray an image of intrinsic

² Nee's work does recognize the role of the state, local corporatism, networks, and the persistence of political power (see also Nee and Lian 1994; Stark and Nee 1989). In my view, however, such a recognition has not altered his primary emphasis on the institutional logic of markets, as Nee put it: "According to market transition theory, the causes giving rise to a change in the mechanisms of stratification in reforming and postcommunist societies are linked inextricably to the expansion of market institutions" (Nee and Matthews 1996, p. 422). Indeed, if other alternative institutional logics were to be conceptualized to have a prominent role, Nee's market transition theory would have lost much of its flavor and would have become indistinguishable from other approaches.

interconnectedness between politics and emerging markets in the course of institutional changes in China, where state-initiated reforms and emerging markets have been two major sources of economic transformation. Both interest politics and markets coevolve in interactions with each other: in some areas, they compete with and constrain each other; in other areas, they mutually reinforce each other; in still others, they adapt to and transform each other in the process. The central idea in the imagery is that neither one can be understood without a careful and substantive understanding of the other.

The premise of the proposed model is based on an insight from institutional theories in sociology and economics that politics and markets are not antithetical and that the state plays a critical role in setting up institutional rules within which markets operate (Campbell, Hollingsworth, and Lindberg 1991; Esping-Andersen 1990; Evans, Rueschemeyer, and Skocpol 1985; Fligstein and Mara-Drita 1996; Wade 1990). Two findings in this literature are especially relevant for my discussion here. First, the expansion of markets is not a self-evolving process. Rather, economic activities are embedded in and constrained by the social context and historical path of change (Granovetter 1985; Hamilton and Biggart 1988). Second, in the coevolution of politics and markets, the state is an active actor, with its own interests and preferences. As a result, changes in economic institutions inevitably reflect a compromise between competing interests and institutions. I find myself in complete agreement with the view of new institutional economics that "the conception of economy is of an evolving, open system in historical time, subject to processes of cumulative causation—instead of approaches to theorizing that focus exclusively on mechanical equilibria" (Hodgson 1994, p. 69).

What are the mechanisms that govern this coevolutionary process? Nee's market transition theory best elaborated one such mechanism—the competition between existing and new economic institutions. The expansion of markets introduces new mechanisms of resource allocation and new ways of organizing economic activities. Redistributive economic institutions, in contrast, rely on the political authorities and favorable regulatory policies to hold on to their advantages in economic production and transactions. So long as market-induced economic activities outperform those of the current (redistributive) institutions, they tend to grow and increase their share in the national economy. In this sense, the emerging market economies compete with and undermine the state socialist redistributive economy.

An equally important mechanism is interest politics in the political arena. Social and economic institutions are associated with organized interests. The course of any institutional change ultimately depends on how various interests play out in the political arena. In comparison with the

new interests associated with emerging markets (Nee 1996), the vested interests in current political and economic institutions have considerable advantage (1) in their access to political authorities and to the processes of policy making and implementation and (2) in their organizing capacity (bureaucratic politics) in protecting and advancing their interests (Shirk 1993; Lieberthal and Lampton 1992). As Parish and Michelson (1996, p. 1043) pointed out: "The type of market compromises that will be arrived at depends as much on bargaining in more participatory political systems as it does on some underlying economic dynamic."

This leads us to consider another mechanism—the central role of the state—in this coevolutionary process. Economic activities, including markets, operate within the institutional rules set by political authorities. From the institutional logic of markets, it is conceivable that the competitive advantages of market-based institutions would force the state and state policies to evolve in favor of market institutions and the new interests associated with these institutions. So long as the state has its own interests in economic growth and national wealth, the comparative institutional advantages, according to this logic, would eventually give the "new interests" an upper hand in "capturing" the state.

However, the state may have its own interests that are not necessarily in congruence with the economic agents in the marketplace. As North and Thomas (1973, p. 8) noted, "the fiscal needs of government may induce the protection of certain property rights which hinder rather than promote growth; therefore we have no guarantee that productive institutional arrangements will emerge." Indeed, concerns about political stability, legitimation, and historical traditions have led to a variety of institutional rules that facilitate as well as constrain markets even in industrialized market economies. Given the historical role of the state in China, and the prevalence of vested interests associated with existing institutions, there is no reason to doubt that the remaking of institutional rules in China's economic transformations will be heavily influenced by the vested interests and the state's own interests.³

Consequently, both markets and politics coevolve in response to each other. Given the constraints of institutional rules and the prevalence of political authorities, market activities take unique forms in China: private enterprises are disguised as "collective" firms; all kinds of economic agents cultivate relations with political authorities and engage in rent-seeking behavior (Parish and Michelson 1996); and economic transactions, even

³ Even the dismantling of the redistributive state would not lead to a break with the past, and vested interests and "old" institutions are likely to continue to exert their influence, as is evident in the experience of other former state socialist societies (Stark 1996; Róna-Tas 1994).

when they operate through the price system, are often brokered by political authorities (Boisot and Child 1996). These demands for and benefits from political authorities reinforce the latter's role in economic life and foster emerging interests that promote the roles of both politics and markets. In this sense, economic activities in the marketplace are molded by politics. On the other hand, the nonstate sector's increasing contribution to state revenue provides positive feedback for the state to adopt policies to encourage market expansion; governmental agencies are gradually changed from redistributors to regulators; and state-owned firms, though constrained by political authorities, are being pushed into market competition. In this process, both politics and markets have been transformed by each other.

I want to point out that the specific processes or mechanisms outlined here have been proposed and elaborated by various scholars in the literature. My intention is not to ignore the distinctive logics behind these theoretical models or to create a false compromise among these arguments that are motivated by different theoretical assumptions. The proposed conceptual model, I hope, serves the purpose to highlight the interconnect-edness among these processes and to insist on understanding them in relation to one another.

How do we assess the outcomes of this coevolutionary process? An important implication of the proposed model (as well as other models that emphasize one or another particular process) is that there is a large area of model indeterminacy with respect to its power of empirical predictions. For instance, in the processes of institutional change, the recognition of the presence of a market institution (e.g., contract, labor market) does not necessarily lead to any predictive power without a substantive understanding of how it interacts with and is constrained by existing institutions. In this light, we can reduce model indeterminacy and improve the power of theory only by substantive institutional analyses of how these causal processes interact and exert their impacts in the specific institutional context, to which I now turn.

Implications for Income Inequality: Hypotheses

Empirically, the debate between market transition theory and those who emphasize the political economy of transformation has centered on the implications of their theoretical arguments for changes in the social stratification order. I now recast these theoretical arguments and their implications for income inequality. Although my discussion recognizes the lineage of these theoretical ideas, my emphasis is on the processes they identified and the ways these processes interact with each other. I take the market processes emphasized in Nee's market transition theory as my starting

point and then consider how other processes may constrain or intertwine with market processes, and their implications for patterns of income determinants.⁴

Inequality between redistributors and producers.—Building on Szelényi's (1978) argument that inequalities "are expressions of the basic conflict of state socialist societies, the conflict between the 'immediate producer' and the 'redistributor'" (p. 77), Nee predicts that market transitions produce fundamental changes in income inequality among social groups. As Nee put it (1996, p. 916), "the growth of market institutions (i.e., labor markets, subcontracting arrangements, capital markets, and business groups) causes a decline in the significance of socialist redistributive power even in the absence of fundamental change in the political order." The empirical implication is that *the advantage of "redistributors" in the acquisition of economic benefits declines relative to that of "producers" in the reform era, compared with the prereform era.*⁵

To the extent that the emergence of markets creates new opportunities outside the redistributive system and entails principles of resource allocation different from the redistributive economy, Nee's proposition is a plausible one. However, one's market position depends on access to opportunities and resources, which, in the Chinese context, is significantly affected by the existing institutional arrangements of redistribution. Those with positional power have advantages in access to both resources and the market-induced opportunities, relative to other social groups: previous "redistributors" may now profit from economic arena in the role of "regulators" in market transactions or as parties of economic transactions (Bian and Logan 1996; Walder 1996). Moreover, the weakening of government intervention into firm-level decisions may have strengthened local authority's power in gaining a larger share of the surplus in profit sharing on the shop floor. Thus, with the advance of markets, some "redistributors" may have lost their benefits, and some social groups may have gained their benefits. But, once we take into consideration these independent processes, it is likely that, *on average, the advantage of "redistributors" in the acquisition of economic benefits may not decline relative to that of "producers" in the reform era, compared with the prereform era.*

An important empirical issue is how to define "redistributors" and "pro-

⁴ My emphasis here is less on adjudicating between competing theoretical arguments than on identifying competing processes and their interactions. Given the state of conceptual and empirical work in the literature, attempts to adjudicate between competing theoretical arguments are premature and may not be productive to advancing our knowledge of the field.

⁵ Because my research design focuses on comparing changes over time, the empirical implications in this section are formulated in the form of a comparison between the prereform and the reform eras.

ducers." Nee recently advocated to define "administrative elites" as "redistributors" and defined "producers" to include, in addition to ordinary workers, "entrepreneurs, managers, and technicians" (1996, p. 916). In this study, I develop two alternative measures of cadre (administrator) status. The first one uses bureaucratic *ranks* of cadres promulgated by the central government. The second one distinguishes cadres in the public sector (government and public organizations) from those in the economic sector (in enterprises). The higher a cadre's bureaucratic rank, the closer he or she is to the redistributive authority than to market activities. Similarly, cadres in the public sector are closer to the political authority than cadres in the economic sector. Therefore, the former is closer to the concept of "redistributors" and the latter to that of "managers." I measure professionals in a similar way to detect the possibility that economic benefits vary with their positions (professional ranks) or their closeness to markets (public versus economic sectors).

Political versus human capital.—Institutional changes are reflected in changes in mechanisms of resource allocation. In a market economy where factors of production are allocated through competitive market transactions, it is argued that human capital plays an important role in determining one's economic rewards (Becker 1964). In contrast, it is often observed that the communist state rewards political loyalty more than competence. Following this logic, Nee (1989) argued that, as the political logic of redistribution is eroded by market mechanisms, *returns to political capital declines and returns to human capital increase in the reform era, compared with the prereform era.*⁶

Now consider other coexisting processes in allocating and rewarding political and human capital in the course of China's economic transformation. First, with regard to political capital, a direct implication can be drawn from the preceding discussion: so long as political authorities play a significant role in bureaucratic recruitment and promotions, persistent and significant returns to political status are logically plausible.

Second, with respect to returns to human capital, one needs to consider multiple and competing processes of allocating human resources in urban China. Nee's argument points to the rise of labor markets in allocating human capital as the source of increasing returns to education. But the state has been and is still actively involved in the allocation of human resources. Even in the reform era, a large proportion of individuals with high human capital (e.g., college education) entered the state sector, especially in government and public organizations (Zhou, Tuma, and Moen

⁶ In the literature, it is conventional to use Communist Party membership as an indicator of political capital, and formal education and work experience as indicators of human capital.

1996, 1997). More importantly, since the early 1980s—before any substantive market reform took place in the urban areas—government policies have emphasized educational credential as one of the most important criteria in obtaining political and positional power (such as party membership and cadre promotion). In this light, *the importance of educational credential may increase, even in the absence of market allocation of human resources.*⁷

The institutional arrangements of work organizations.—In state socialist China, work organizations have been major institutions of redistribution. An important contribution by scholars who study social stratification processes in China is the recognition and theoretical arguments about the centrality of work organizations in the socialist redistributive system (Bian 1994; Lin and Bian 1991; Walder 1986, 1992; Whyte and Parish 1984; Zhou et al. 1996, 1997). These studies identified a hierarchical order among types of work organizations whose economic benefits vary systematically with state policies and their property rights relationships to the state. Work organizations in the state sector (government agencies, public organizations, and state-owned firms) benefit more from redistribution and state policies than those in the semistate and nonstate sectors (collective firms, hybrid, and private firms).

How would the expansion of markets affect the institutional arrangements of work organizations? According to the logic of market transition theory (Nee 1992), one would argue that firms that are closer to market transactions are more likely to experience major changes in patterns of income determinants, relative to those organizations in the state sector. Because political control varies significantly across types of work organizations, those who are less controlled by the state are more likely to break away from the redistribution system. Thus, with the advance of market processes, we expect that organizational hierarchies associated with the redistributive economy would change, so would the distribution of economic resources associated with types of work organizations. That is, *in the reform era, patterns of economic benefits associated with types of work organizations should favor those organizations (collective and hybrid/private firms) that are closer to market transactions and less controlled by the redistributive system.*

However, there are important competing processes that countervail the

⁷ There is no intrinsic principle of state socialist redistribution that dictates low returns to education. As Konrad and Szelenyi (1979) pointed out, socialist planning systems also value education and knowledge and incorporate intellectuals into political elites. A recent comparative study of redistributive patterns in the USSR and China reveals that returns to education were considerably higher in the USSR than in China (Zhou and Suhomlinova 1999).

market-induced processes identified above. Work organizations in the state sector serve legitimation for the political order and the interests of the central authority (Walder 1986). Relative to the new interests emerging from market expansion, the vested interests associated with existing institutions are better organized and have better access to political authorities. Institution-based bureaucratic bargaining often leads to those policies and regulations that protect the vested interests associated with the state sector (Oberschall 1996; Parish and Michelson 1996; Zhou et al. 1997). Hence, these considerations point to state intervention in regulation or resource transfer in favor of those in the state sector, producing institutional persistence. In this view, *the politics of vested interests embedded in existing institutions will lead to the persistence of income inequality based on existing organizational hierarchies in the reform era.*

DATA

The empirical analyses are based on life histories of a sample of 5,000 residents drawn from 20 cities in China, which interviewers collected in the summers of 1993 and 1994. We collected retrospective information on respondents' locations in type of work organizations, occupations, and residential locations, as well as their education and political status (party membership) over time.

We selected six provinces (Hebei, Heilongjiang, Gansu, Guangdong, Jiangsu, and Sichuan), each of which represents a conventional geographic region in China. In each province, we chose the capital city to represent large cities (population over 1 million). We also randomly selected a medium-sized city (population between 200,000 and 1 million) and a small city (population under 200,000) based on the 1990 *Yearbook of Chinese Cities* (SSB 1990a). The sample size in each city was proportional to the population in that size of city in that province. We also included Beijing, China's capital, and Shanghai, its largest industrial city. These 20 cities cover a variety of geographic locations and different types of urban economies.

We chose a stratified random sample of each city's residents. In China, a metropolitan city is composed of residential districts. Each district is composed of residential "streets," and each residential street is organized into residential blocks (*juweihui*). We selected residential blocks in each city using a systematic sampling scheme. That is, we selected every N th residential block based on the official residential statistics. We used an analogous sampling scheme to select households in each residential block. Finally, we randomly selected a member of the household between ages 25 and 65 to be interviewed.

We collected income information for the following years: 1955, 1960, 1965, 1975, 1978, 1984, 1987, 1991, 1992, 1993, and 1994. Most of the selected years were associated with important historical events. For instance, 1960 was the year of a severe economic disaster, and 1965 was the year on the eve of the Cultural Revolution. We selected these years for several reasons. First, income distribution in years of historical significance helps us identify important characteristics of the redistributive system. Second, the salience of these years in one's life experience may assist the respondent to recall information more accurately.

The data yield an urban labor force sample whose composition varies over time. An individual enters my analyses when she enters the labor force and leaves the analyses when she exits from the labor force (e.g., retires). I also exclude the income records for those in the rural labor force in a particular year, because incomes in rural and urban areas are not comparable.

Recall errors are inevitable in retrospectively collected data, especially with respect to income. However, an advantage in collecting income information in China is that there were few wage changes in the prereform era. In the reform era, income changes were more frequent, but these events are more recent and easier to recall. To assess the quality of the income data, I conducted a comparison between the information collected in our sample and that based on official statistics over the years. Specifically, I compared the average income of our sample base on respondents' self reports in those selected years and the average income of our sample using official statistics of the average income in those provinces where the respondents were drawn.⁸ Figure 1 displays the patterns between the two.

As figure 1 indicates, between 1955 and 1975, average income in our data is close to that based on official statistics. The differences between the two increased after 1978, with the average income in our sample being higher than the official statistics. But these differences are not substantial relative to the income level in these specific years. More importantly, the differences between the sample and official statistics show a relatively stable trend. I believe that the higher average income in our sample in recent years is mainly due to the fact that our sample has a larger proportion of residents from large cities (Beijing, Shanghai, and capital cities of the provinces). Average income levels in large cities are higher than other cities, especially in the reform era. The official statistic is based on the average income of the urban areas in an entire province, which should

⁸ It is desirable to use more detailed information (e.g., age groups, city-level information) for such a comparison. Unfortunately, there is no systematic information over time in official statistics.

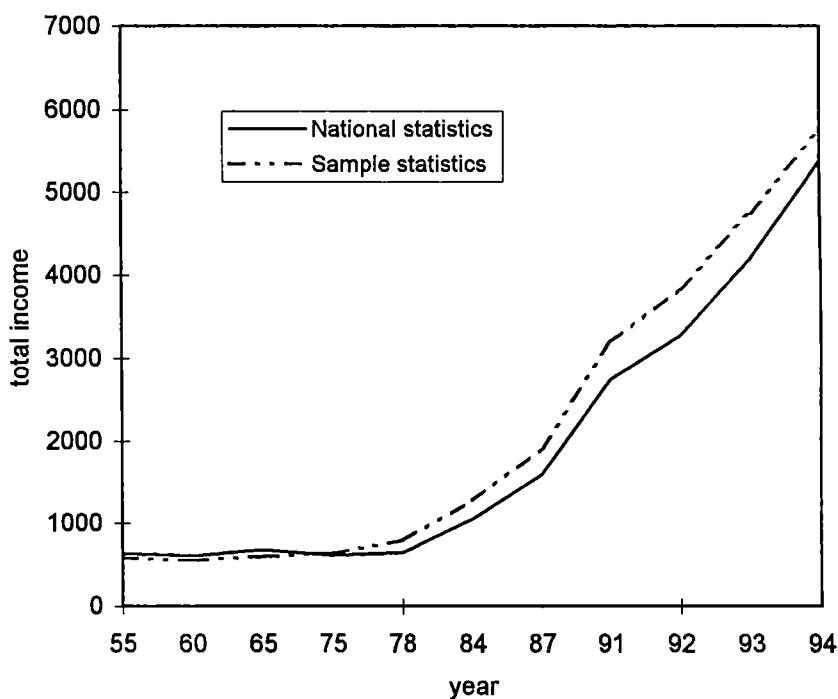


FIG. 1.—Comparison of national and sample statistics in income

have lower average income. Thus, the source of the observed discrepancies can be captured by modeling city locations in statistical analysis.

I conducted extensive data checking and deleted those records where the reported income was logically suspect. Cases with missing values in income or the covariates were excluded from the analyses. All in all, the number of deleted cases is relatively small (about 5%). The actual number of cases used in the analyses is 4,730.

VARIABLES

Dependent Variable

Personal income.—I use the logarithm of total personal income (Chinese yuan/month) as the dependent variable. We collected information on respondents' basic income, bonus, and income from other sources. Thus, total income reflects both rewards from the current job and other sources based on one's skill or position. Income is adjusted for inflation using provincial-level information (SSB 1990b).

Independent Variables

As I noted before, operationalization of key theoretical concepts is crucial to provide a common ground to compare research findings across studies. The principle I adopted is, wherever possible, to use generic categories whose meanings are widely shared to avoid artificial ambiguities in interpretation.

Gender.—I use a dummy variable (female = 1) to examine gender-based variations in the redistribution of economic rewards.

Age and age².—I use the first- and second-order effects of age to measure the effect of work experience or seniority, as part of human capital.

Education.—Formal education is a conventional measure of human capital. I distinguish the following educational levels: (1) illiterate or elementary (the reference category); (2) junior high; (3) senior high (including *Zhongzhuan*); and (4) college (including *Dazhuan*).

Party membership.—I use a dummy variable (party membership = 1) to indicate Communist Party membership as political capital.⁹

Occupation.—Occupational groups are closely related to positional power of social groups. For this reason, I develop two alternative measures of cadre status. In the first one, I distinguish the following occupational categories: high-rank cadre, low-rank cadre, high-rank professional, low-rank professional, clerk, service worker, skilled worker, and unskilled worker (the reference category). The Chinese bureaucratic hierarchy has mainly four levels: *bu* (ministry), *ju* (department), *chu* (division), and *ke* (section). I classify those holding ranks at or above *chu* level as high-rank cadres and those at or below *ke* level as low-rank cadres. In the Chinese professional system, there are senior engineer, engineer, assistant engineer, and technician levels (or equivalent levels in other professional occupations). I classify those at or above engineer level as high-rank professionals and those at or below assistant engineer level as low-rank professionals.

In the second specification, I distinguish cadres in government and public organizations (the public sector) from those in production/service organizations (the economic sector). Similarly, I distinguish professionals in public and economic sectors. Because of the small number of cases, I do not differentiate ranks among cadres and professionals in this specification. Other occupational categories remain the same.

Work organization.—I distinguish the following types of work organizations:

⁹ Because samples from two provinces (Jiangsu and Guangdong) do not have information on party membership, I examine the effect of party membership in a separate analysis, excluding respondents from these provinces.

1. Governmental agencies include ministries, commissions, bureaus, and offices at various levels of the Communist Party and central and local governments.
2. Public organizations, in Chinese terminology, are nonprofit organizations in the public domain. They include educational and research institutions and organizations in the medical, publishing, broadcasting, and entertainment sectors. Although they are not the administrative organs of the state, most of these organizations are affiliated with the state or local governments through financial and organizational linkages.
3. Central government-owned firms. Among state-owned firms, I distinguish two categories: the central government-owned firms and local government-owned firms. Included in the first category are those work units in manufacturing, processing and other production firms, and those in service sectors that are directly owned by the central or the provincial government.
4. Local government-owned firms. I include firms that are state-owned but are managed by local (city or districts within cities) government in this category. These firms benefit from redistribution associated with the state sector but closer to local authorities.
5. Collective firms. Organizations in this category are not directly under the administration or financial support of the planning economy. Often they are sponsored by local governments (such as district/county government or residential offices). This type of organization has the least redistributive benefits but is closer to market transactions and less regulated by the government than state firms.
6. Private/hybrid firms. This type of firm includes private entrepreneurs, firms with mixed property rights, such as partly collective-owned and partly private-owned, or joint ventures between state-owned firms and foreign firms.

Control variables.—I include a set of dummy variables to indicate respondents' residential location (city) in all analyses to control for city-specific variations in income. Table 1 reports the attributes of the covariates in selected years.¹⁰

¹⁰ The distribution of some categories in the earlier years (e.g., party membership and high-rank cadres in 1960) appears to be high for the particular age cohort (18–32) in our sample. Because of the lack of official statistics, I am unable to assess this issue accurately. There are several plausible explanations for the observed patterns. First, because the sample size in the earlier years is small, the distribution of certain attributes is likely to be sensitive to even random sampling variation. Second, our sample has a larger proportion of residents from large cities who had better access to party membership and promotion than the national average.

TABLE 1

DESCRIPTIVE STATISTICS OF THE COVARIATES, SELECTED YEARS

Covariates	1960	1965	1975	1978	1984	1987	1991	1993
Age ^a	25.3	28.3	32.7	33.7	35.5	36.5	38.2	39.3
Female47	.45	.48	.47	.44	.43	.41	.41
Party membership ^b19	.20	.20	.20	.21	.23	.23	.23
Education:								
Elementary or below50	.41	.30	.27	.20	.16	.13	.12
Junior high21	.24	.34	.35	.34	.34	.33	.33
Senior high17	.21	.23	.27	.34	.35	.36	.36
College08	.11	.10	.09	.11	.13	.17	.18
Occupation:								
High-rank cadre02	.02	.02	.02	.03	.03	.03	.03
Low-rank cadre11	.13	.12	.12	.12	.12	.13	.13
High-rank professional02	.04	.03	.04	.05	.06	.07	.08
Low-rank professional16	.19	.17	.15	.14	.13	.14	.14
Clerk04	.04	.04	.04	.05	.05	.06	.07
Service worker09	.09	.09	.10	.12	.12	.12	.12
Skilled worker32	.30	.32	.32	.30	.29	.27	.26
Unskilled worker22	.21	.21	.21	.19	.17	.15	.14
Private entrepreneurs01	.02	.03	.04
Cadre in public sector06	.07	.06	.06	.05	.06	.06	.07
Cadre in economic sector07	.07	.08	.08	.09	.09	.09	.09
Professional in public sector10	.12	.09	.09	.08	.09	.09	.10
Professional in economic sector08	.10	.11	.10	.10	.11	.12	.12
Type of workplace:								
Government agency11	.11	.09	.09	.08	.09	.10	.10
Public organization13	.14	.12	.11	.12	.12	.12	.13
Central government firm32	.30	.30	.29	.28	.28	.26	.26
Local government firm20	.19	.21	.22	.22	.22	.22	.21
Collective firm21	.24	.26	.26	.27	.25	.23	.22
Private/hybrid firms03	.02	.03	.03	.04	.05	.07	.09
N	805	1,164	2,174	2,705	3,379	3,605	3,699	3,678

^a Age refers to the mean age in the sample. All other entries refer to proportions in that category.^b Samples from two provinces (about 20% of the total sample) have no information on party membership. The percentage reported here is calculated after excluding these samples.

MODELS AND METHODS

Multivariate Regression Model for Cross-Sectional Data Analysis

As a preliminary analysis, I use the conventional multivariate regression model to examine income determinants in selected years. This set of analyses aims at detecting variations in income determinants over different historical contexts.

The analysis of cross-sectional data is limited in addressing some important theoretical issues. For instance, "Communist Party" membership is often used as an indicator of political capital. But, the Communist Party may selectively recruit those who had more ability into the party. If this is indeed so, party membership may reflect the compound effects of both political loyalty and ability. However, in a cross-sectional research design, it is impossible to distinguish the two effects.

A Mixed Model for Panel Data Analysis

My main analytic focus is to model the panel data for the analysis of changes in income determinants in two historical periods: the prereform era and the reform era. Using multiple observations for each subject over time, models based on panel data can effectively control for unobserved individual attributes such as "ability" (Hsiao 1986; Judge et al. 1985). This set of analyses aims at examining income determinants of theoretical importance after controlling for important but latent individual attributes as well as random fluctuations across cities and years.

As I noted before, interpretation of empirical results is often contentious because of disagreement over the operationalization of theoretical concepts. One advantage of the panel data analysis in this study is that, by comparing changes across periods using the same measurements of key concepts, the consistency of the measures over time partly alleviates potential impreciseness in measurements. The estimated pattern of stratification in the prereform era also allows an empirical assessment as to whether the operationalization of the concepts and model specification have captured the redistributive patterns discussed in the literature.

An important issue in panel data analysis is to model correlations among repeated measures within a subject. Since panel data analysis makes use of information on a series of observations for each subject (e.g., income in several years for a respondent), observations within a subject are "clustered" and their error terms are correlated.

In our data, there is a second source of clustering. Our data are drawn from 20 cities, and there is considerable heterogeneity across cities. Individuals from the same city are influenced by similar economic conditions (inflation, industrial structures, economic development, etc.). Also, respon-

dents' income in a particular year may be affected by broader economic conditions or political events in that year. Therefore subjects in a specific year may also be "clustered." The sources of both within- and cross-subject "clustering" violate the independently and identically distributed (i.i.d.) assumption in the OLS estimation, and one needs to explicitly model the covariance structures that deal with these sources. For this purpose, I propose a mixed model for the panel data analysis of income. The general form of a mixed model can be expressed as follows:

$$Y = X\beta + Z\gamma + \epsilon, \quad (1)$$

where X is a set of covariates treated as having fixed effects, Z the set of covariates having random effects. The variance of Y , $\text{var}(Y)$, is

$$\text{var}(Y) = ZGZ' + R. \quad (2)$$

The first source of clustering, repeated measures within a subject, is mainly related to modeling the R vector in equation (2). Conventionally, it is sensible to consider some kind of time-series covariance structure. However, in our data, the observation points within a subject (the selected years of income) are not equally spaced. As a result, conventional time-series covariance structures (e.g., autoregressive error process [AR] 1) are inappropriate because of their equal-spacing assumption.

As Littell et al. (1996, p. 127) pointed out, many spatial covariance structures can be generalized to model covariance structures in which the correlations decline as a function of time. The spatial power function is a direct generalization of the AR(1) structure but allows unequal spacing among observations within a subject. It models the covariance between two measurements at times T_1 and T_2 as

$$\text{cov}(y_{i1}, y_{i2}) = \sigma^2 \rho^{|t_1 - t_2|}, \quad (3)$$

where ρ is an autoregressive parameter assumed to satisfy $|\rho| < 1$ and σ^2 is an overall variance.

In addition, it is likely that the observations have measurement errors. One can model the measurement errors by allowing an additional variance parameter in R to capture measurement errors across subjects. I further specify $\text{Var}(\epsilon_i) = \sigma^2 + \tau^2$, where τ^2 captures measurement errors in a subject. Thus, the R vector is specified as

$$R = \tau^2 I + \sigma^2 \rho^{|t_1 - t_2|}. \quad (4)$$

For the second source of clustering—correlations among subjects within city localities and in historical years—I propose to model city localities and years as having random effects. That is, I allow the effects of city localities and the selected years to vary randomly to account for city- and year-specific factors that are not included in the model. To accomplish

this, I include a set of indicator variables in the Z vector. The corresponding G variance/covariance structure is specified as variance-component structure, with each city (year) having its own variance parameter but uncorrelated across cities (years). My primary interest is related to the fixed effects of the covariates in X , after controlling for the effects of city- and year-specific factors.

To assess changes in the social stratification order over time, I estimate the interaction effects between the sets of covariates in X and a dummy variable indicating the reform era in urban China. In this formulation, the main effect of a covariate indicates its overall effect on income in urban China. A *statistically significant* effect of the interaction term for that covariate indicates a “statistically discernible” change in the effect of that covariate *in the reform era*. The direction and magnitude of such a change can be interpreted based on the parameter estimates of that interaction term.

To sum up, I specify a mixed model as follows:

$$\ln(Y) = \alpha + X_{it}\beta + X_{itp}\lambda_p + Z_{city}\gamma + Z_{year}\delta + \epsilon_{it}, \quad (5)$$

where X_{it} refers to the covariates of theoretical interest that vary over time and across individuals. In my model specification, I treat this set of covariates as having fixed effects. β are the corresponding parameter estimates. The λ vector is the set of the estimates of the interaction between X and the period indicator, p . Z vector contains the set of indicator variables for city localities and the selected years for which income information was collected, and γ and δ the corresponding estimates of their respective random effects. In addition, equation (5) also estimates v_{city} and μ_{year} , the covariance components associated with city localities and the selected years respectively. R follows the specification in equation (4), and ϵ_{it} is the structure of the error terms for the specific observations. It is assumed that ϵ_{it} follows an i.i.d. distribution, after controlling for the variance/covariance structures specified above. The appendix provides a detailed discussion of model specification in this study.

Because of the unequal-spaced observation points and potential missing information involved in the panel data, maximum likelihood estimation produces biased estimates of the random parameters. The restricted maximum likelihood (REML) procedure can deal with this problem and is adopted in the model estimation in this study.

RESULTS

Determinants of Income: A Cross-Sectional Data Analysis

I first examine patterns of income determinants in selected years using a multivariate regression model. My purpose is twofold: First, the results

in this set of analyses help detect patterns of change in income determinants over time. Second, my analyses establish a basis for comparison with results in previous studies and in the panel data analysis in later sections.

Table 2 reports the parameter estimates of the covariates in selected years. The estimated models also included a set of indicator variables for respondents' city localities. My interest is in patterns of resource allocation among social groups, while controlling for cross-city variations in overall income levels.

As indicated at the bottom of table 2, sample sizes vary across these selected years, reflecting changing compositions of the sample over time. The R^2 indicates that the model specified for these selected years has a reasonably good explanatory power. This model accounts for 24%–44% of the income variation in various years.¹¹

We can detect roughly two income regimes over time and a transitional period in between. The first income regime is associated with the Mao era, 1949–77. During this period, there was a general trend of equalization in income. Across the three years under this regime (1960, 1965, and 1975), there was a trend of decreasing gender inequality; differences in occupational status (e.g., returns to cadre and professional status) were also narrowing. There was some reduction of income differences among types of work organizations, but the patterns were less systematic.¹²

Patterns of income determinants in 1978 and 1984 reflected a transitional period between the prereform and reform eras. By 1978, the new leadership consolidated its power. But little institutional change was introduced in urban areas. As a rectification of the radical policies of the Cultural Revolution, the government adopted a series of wage increases, which were reflected in increases in returns to cadre and professional status in 1978 (see table 2). In some aspects, patterns of income determinants in 1978 resembled a return to the pre-Cultural Revolution period.

In the early 1980s, the central government experimented with decentralization in allocation of resources in state-owned firms. The most immediate effects of these policies on income were that managers in firms had the discretion to decide bonuses outside the state-set wage system. These

¹¹ The high R^2 is partly due to the inclusion of the city indicator variables, which captures substantial variations in income across cities.

¹² Because of the retrospective nature of the sample, variations in income determinants across the years partly reflect the evolving composition of the sample. The patterns in the earlier years may reflect a younger population. I replicated the analyses (and those reported in table 3 below) using only 35% of the sample (those who entered the labor force before 1966). The patterns in the replication study are similar to those reported in table 2, indicating that the reported patterns are robust to variations in sample composition over time.

TABLE 2

OLS ESTIMATES OF THE DETERMINANTS OF INCOME, SELECTED YEARS

Covariates	1960	1965	1975	1978	1984	1987	1991	1993
Intercept	1.97***	2.28***	2.32***	2.43***	3.80***	3.75***	3.34***	3.33***
Female	-.29***	-.24***	-.18***	-.16***	-.14***	-.19***	-.18***	-.18***
Age09*	.07**	.06***	.06***	.01†	.02***	.04***	.04***
Age ² /100	-.11	-.07†	-.06***	-.07***	-.01	-.02**	-.04***	-.04***
Education:								
Junior high04	.04	.06**	.04	.05†	.07**	.09**	.11***
Senior high11*	.06†	.05†	.03	.02	.10***	.16***	.18***
College18**	.13**	.15***	.07†	.03	.16***	.16***	.17***
Occupation:								
High-rank cadre42***	.35***	.29***	.40***	.28***	.21***	.21***	.34***
Low-rank cadre16**	.07	.08*	.16***	.09*	.06†	.10**	.21***
High-rank professional24*	.25***	.13*	.18**	.12*	.14**	.13**	.20***
Low-rank professional20***	.15***	.05	.06†	.02	.03	.07*	.15***
Clerk04	.08	-.02	-.03	.02	-.01	.01	.07†
Service worker00	.03	.01	.05	.04	.06†	.05†	.07*
Skilled worker07	.07*	.02	.06*	.00	-.01	-.01	.03
Workplace:								
Government agency19**	.14**	.09*	.08*	.07†	.07†	.10**	.17***
Public organization21***	.11**	.10**	.09**	.05	.08*	.13***	.16***
Central government firm21***	.12***	.15***	.16***	.12***	.11***	.11***	.17***
Local government firm16***	.11**	.12***	.11***	.07**	.10***	.10***	.11***
Private/hybrid forms18*	.07	.09†	.21***	.30***	.54***	.57***	.56***
R ²42	.40	.36	.30	.24	.29	.40	.44
N	779	1,132	2,134	2,624	3,255	3,478	3,587	3,623

NOTE.—“Elementary school education or below” is the reference category for education, “unskilled workers” for “occupation,” and “collective firm” for “type of work organization.” The models estimated above also included 19 dummy variables for each city included in the analysis.

† $P < .10$.

* $P < .05$.

** $P < .01$.

*** $P < .001$.

changes were reflected in the patterns of income determinants in 1984, where employees in government and public organizations, on average, had no advantages, other things being equal.

Patterns of income determinants in 1987, 1991, and 1993 showed evidence of a new income regime. Overall, the effects of the covariates showed a consistent trend during this period. In particular, we observed increasing returns to education, as well as to cadre and professional positions. Types of work organizations remained important during this period, and their effects appeared to be increasing.

To sum up, the results reported here are broadly consistent with both documented historical changes of the redistributive economy in urban China (Whyte and Parish 1984) and findings in other studies. The general patterns of income variations reflect two distinctive regimes. In the Mao era, 1949–77, variations in income showed a decreasing trend, with the declining importance of education and occupation, among others. This was clearly attributable to the “destratification” state policies promulgated during this period (Parish 1984). However, in the era of urban reform, we observe a trend of increasing inequality, as indicated by the increasing importance of education, occupation, and work organizations.

Changes in the transitional period between 1978 and 1984 were largely induced by state policies rather than by the reform of the redistributive institutional arrangements (with the exception of the growth of the hybrid and private firms).¹³ During this period, the government experimented with a series of reform measures within the framework of existing institutions, especially in the area of wage policies. These changes did not reflect the traditional logic of redistribution or that of markets (Walder 1987). For the purpose of examining institutional changes in urban China, in my view, these two years were closer to the prereform era than to the reform era.

Explaining Variations in Income: A Mixed Model for Panel Data

I now turn to panel data analysis. Based on the findings in the previous analyses, I used the year 1985 to divide the prereform and reform eras in urban China.¹⁴ Given the specific years for which we collected income

¹³ In all analyses reported in this study, I combined “private entrepreneurs” and “those working in private and hybrid firms” into a single group (labeled “hybrid firm” in the tables) because of the considerable overlap between the two categories. My exploratory analysis shows that including both categories in the model does not lead to significant changes in the results.

¹⁴ 1985 marked the official beginning of the urban reform and of the second income regime. It also witnessed a state-initiated wage reform that was based on principles (positions and ranks) substantially different from earlier wage systems. However, it could be argued that there were initial reform efforts in urban areas since the early

information, 1955–84 and 1987–94 were the time spans of the two eras covered in our data. To test the hypotheses about patterns of change across the two eras, I compare the “main effects” of the covariates and the “interaction effects” between these covariates and the period dummy variable, p ($p = 1$ if year > 1985 ; $p = 0$, otherwise).

Table 3 reports the parameter estimates of two models, using the two alternative specifications of cadre/professional status.¹⁵ In model 1, cadres and professionals were specified into “high-rank” and “low-rank” groups respectively. In model 2, cadres and professionals were differentiated between those in the public sector (government and public organizations) and those in the economic sector (enterprises).

Columns 1 and 2 of table 3 show the main effects and interaction effects of the covariates in model 1. The intercept and the associated interaction effect indicate the “overall” income level and changes in the overall income level in the reform era. The positive and significant effect of the interaction term shows that there were statistically significant and substantial increase ($\exp[0.83]-1 = 129\%$) in the overall income in the reform era. However, my interest is in changes associated with specific positions and groups across the two periods.

Gender.—In the prereform era, a female employee earned 84% ($\exp[-.173] = .84$) that of a male employee’s income. In the reform era, there was a negative “female” effect, indicating that female employees’ earning decreased during this period relative to male employees. But, this interaction effect was not statistically significant. That is, we cannot reject the null hypothesis that the observed difference in gender effects across the two periods is due to sampling variations.

Work experience and education.—In the literature, it is conventional to use work experience and formal education to measure human capital. Work experience, as measured by the first- and second-order effects of

1980s. Given the transitional characteristics of income determinants in 1978 and 1984, it is desirable to analyze these two years separately. But due to the small number of data points in this short time span, a separate panel analysis is not feasible. In my preliminary analyses, I explored two alternative specifications of the two eras. The first one was to divide the periods between 1955–78 and 1984–94. The results based on this specification showed *less significant changes* in parameter estimates across the two eras than those reported in the text. In the second specification, I omitted information in 1984 and divided the periods between 1955–78 and 1987–94 on the ground that income determinants in 1984 appeared to be an anomaly. The results showed similar patterns as those reported in the text.

¹⁵ The models also estimated coefficients for the random effects of city and year indicator variables. Because of space limitations and because my interests are in the fixed effects of the theoretically related covariates, I do not report this set of parameter estimates.

TABLE 3
PARAMETER ESTIMATES OF THE COVARIATES IN THE FULL MIXED MODEL

COVARIATES	MODEL 1		MODEL 2	
	Main Effects	Interaction Effects ^a	Main Effects	Interaction Effects ^a
Intercept	3.037***	.830***	3.019***	.812***
Female	-.173***	-.020	-.174***	-.021
Age064***	-.023***	.063***	-.023***
Age ² /100	-.070***	.027***	-.069***	.027***
Education:				
Junior high036*	.049*	.039*	.053**
Senior high073***	.085***	.077***	.091***
College104***	.106***	.114***	.118***
Occupation:				
High-rank cadre239***	.005
Low-rank cadre116***	.002
High-rank professional110***	.050
Low-rank professional056**	.028
Cadre in public sector193***	-.046
Cadre in economic sector103***	.018
Professional in public sector088**	.028
Professional in economic sector055**	.033
Clerk026	.011	.034	.005
Service worker021	.014	.023	.013
Skilled worker019	-.016	.018	-.016
Work organizations:				
Government agency099***	.027	.063*	.054†
Public organization086***	.040†	.058*	.054†
Central government firm142***	.006	.141***	.006
Local government firm075***	.023	.075***	.023
Private/hybrid firms230***	.232***	.230***	.233***
G-covariance structure:				
ν_{city}043**	.017**	.043**	.017**
μ_{year}016*016*	...
R-covariance structure:				
σ210***211***	...
ρ	-.00007***00006***	...
ϵ024***023***	...
<i>N</i>	4,730		4,730	
<i>N</i> × <i>T</i>	27,392		27,392	

NOTE.—“Elementary school education or below” is the reference category for education, “unskilled workers” for “occupation,” and “collective firm” for “type of work organization.” Both models also estimated a set of coefficients for the random effects of cities and the selected years.

^a The interaction effects indicate changes in the effects of the particular covariates in the 1987–94 period.

† $P < .10$.

* $P < .05$.

** $P < .01$.

*** $P < .001$.

age showed significant and varying effects across the two periods. In the prereform era, the effect of work experience showed a familiar inverted-U shape. In the second period, however, we observed changes in both the first-order and the second-order effects of age. This implies that there was a sharp decline in returns to work experience in the reform era. Because the sample evolved over time, the age effects may be sensitive to variations in sample composition. For this reason, I will not give substantive interpretation of age effects in this study.

With regard to formal education, there were significant returns to all educational levels in the prereform era, and these returns increased in the reform era. As the main effects indicated, in the prereform era, a college education increased one's income by 11%, and a senior high education by 8%, compared with those who had elementary or no education (the reference category). The estimates of the interaction effects show that there were significant increases in returns to these educational levels in the reform era. A college degree in the reform era had 23% higher income ($\exp[.104 + .106] = 23\%$), and a senior high school graduate had 17% higher income, compared with those who had elementary or no education. As noted before, this finding may reflect confounded effects of market forces and state policies.¹⁶

Occupation.—The main issues of contention in the debate are related to the role of positional power in the transformation processes. The parameter estimates for occupational groups show returns to various occupational statuses. The "main effects" indicated that, in the prereform era, high-rank cadres had the highest income—27% higher than unskilled workers (the reference category), other things being equal. Professionals (both high-rank and low-rank) and low-rank cadres also had significantly higher income, but the magnitudes appeared to be substantively smaller than that for high-rank cadres. There were no statistically discernible differences among other occupational groups (clerks, service workers, and skilled workers), as compared with unskilled workers.

Market transition theory predicts that returns to the cadre status decline relative to that of producers in the marketization processes. The interaction effects for occupational groups show that there were no statistically discernible changes for these occupational groups in the reform era. The magnitudes for the professionals appeared to be larger than others. But we cannot reject the null hypothesis that these differences are due to statistical variability rather than substantive changes.

¹⁶ As Mincer (1974) showed, the estimates of educational effects (school years) may be biased downward when age is included in a typical wage model. Thus we need to be cautious in interpreting the magnitudes of returns to education in this analysis. Because of the use of panel data in this study, the comparison of *changes* in returns to education between the two eras is less likely to be affected by this problem.

Type of work organization.—Organizational hierarchy was the center of the redistributive economy in urban China. As the results of the “main effects” show, in the prereform era, such an organizational hierarchy was indeed present. Compared with the reference category of collective firms, employees in central government-owned firms enjoyed the highest income, other things being equal. Those in governmental agencies and public organizations, as well as those in local government-owned firms, also had advantages. This pattern is consistent with the organizational hierarchy of the redistributive economy and prevalent state policies documented in the literature. The main exception is the significant and higher income for employees in the private/hybrid firms. This resulted from the rise of a private sector in urban economies since 1978, as our cross-sectional analyses (see table 2) indicated.

To what extent has this organizational hierarchy changed in the reform era? According to the results from the “interaction effects,” not much. Except for the significant and positive effect for private/hybrid firms, the economic benefits associated with types of work organizations have not shown statistically discernible changes. Moreover, collective firms had not benefited from market activities, relative to other types of work organizations, despite the fact that they were closer to market activities than those workplaces in the state sector. In other words, there is no strong evidence that the institutional arrangements of work organizations and the associated patterns of income distribution in urban China have changed significantly.

Party membership.—In a separate analysis similar to that reported in table 3 but excluding samples with no information on party membership (samples from Guangdong and Jiangsu Provinces), I estimated that the net return to party membership is about 6% ($\beta = .058$, $P < .001$), and there is no significant change in the reform era, other things being equal. That is, there is no evidence of “declining in significance” of political capital across the two periods. I note that because a large proportion of employees in the urban labor force are party members (over 20% on the national average), this political status is widely diffused, raising questions about whether it is still a useful indicator of “political capital” in urban areas.

Findings based on the second measure of cadre status.—Given the critical role of cadre/professional status in the theoretical debate, model 2 adopted alternative measures that distinguish cadre and professionals in two sectors. According to the logic of market transition theory, managers and professionals in economic sectors would benefit more from market transactions relative to those in the public sector who were more regulated by state policies.

As the parameter estimates in model 2 show, in the prereform era, cad-

res in the public sector had the highest income, cadres in the economic sector and professionals in both public and economic sectors also had significant and higher income than unskilled workers (the reference category), but their advantages were less salient. However, there were no statistically discernible changes in these patterns in the reform era, as indicated by their corresponding interaction effects. The interaction effect did point to a decrease in income for cadres in the public sector. However, this coefficient was neither substantial (in magnitude) nor statistically significant to warrant a different conclusion.

The effects of government and public organizations were smaller in model 2 than in model 1. This was because the specifications of cadres and professionals were based on their locations in types of work organizations. Clearly, a large proportion of cadres and professionals were concentrated in government and public organizations. In the reform era, government and public organizations had marginally significant ($P < .10$) and higher income than before. Other types of work organizations had no significant changes, except for income increase for employees in hybrid/private firms.

Variance components.—Parameter estimates of variance components in the G -vector show significant “clustering” effects associated with city localities and the selected years. The estimates of the parameters in the R vector also show evidence of autoregressive correlation among repeated measures within a subject.

Summary.—The main evidence for the impacts of market expansion, as consistent with market transition theory, is the significant and higher returns to those working in private and hybrid firms, whose economic activities were more governed by market transactions than by the redistributive economy. The increasing returns to education are also consistent with market transition arguments.

On the other hand, there is no evidence of “decline in significance” in returns to positional power, as measured by cadre/professional status in two alternative specifications. Besides the significant role of private/hybrid firms, there is no evidence of substantial changes in the organizational hierarchy in urban economies.¹⁷

¹⁷ I also conducted statistical analyses that compared changes in income determinants over time among the inland region (Gansu, Heilongjiang, Sichuan, Hebei), coastal regions (Jiangsu, Guangdong), Shanghai, and Beijing. There were noticeable regional variations in the effects of the covariates in this set of analyses. But these results are largely consistent with the overall patterns reported in table 3. Because of the smaller number of cases in each region, the statistical power is relatively weak. I did not report this set of analyses due to space limitations.

Changes in Income Determinants in the 1990s: An Alternative Baseline

In the 1990s, economic reform in urban China has accelerated. Therefore, to detect more recent changes in the processes of marketization, one may use an alternative baseline to compare changes in patterns of income determinants between 1955–87 and 1991–94.¹⁸ I estimated the changes in income determinants in the 1990s, using this alternative baseline. Table 4 reports the parameter estimates from this set of analyses.

Returns to education.—As table 4 shows, in both models 1 and 2, changes in returns to education in the 1990s are significant for all educational levels, but the magnitudes of increases for senior high and college are smaller in the 1990s, as compared with findings in table 3. These results seem to suggest, after more rapid changes in the 1980s, increases in returns to education slowed down in the 1990–94 period.

Returns to cadre and professional status.—In model 1, low-rank cadres and professionals had more significant improvement in their income in the 1990s than before. This finding suggests that the advancement of markets led to improved economic benefits for those cadres and professionals who were more distant from the redistributive center. But the advantages of the high-rank cadres have not changed relative to the large segments of “producers”—skilled and unskilled workers.

Findings in model 2 show that cadres and professionals in both public and economic sectors gained in income in the 1990s compared with the earlier period. Cadres in the public sector gained most, and professionals in the economic sector had the second largest gain. These results reflected confounding processes that reward both those closer to the political authority and those closer to market transactions.

Returns to work organizations.—In model 1, both government agencies and public organizations had the most significant gains in the 1990s, compared with collective firms. Hybrid firms had the largest gain, among all types of work organizations. However, their magnitude was smaller than those reported in table 3. In model 2, there were no significant changes in the hierarchical order of work organizations between the two periods.

Overall, more dramatic changes in the 1990s, with both the expansion of markets and the weakening of the central government, did not lead to a significant alteration of the patterns of income determinants as compared with the findings in table 3. These findings reinforce the view that institutional changes involve multifaceted processes that produce institutional changes in some areas but institutional persistence in others.

¹⁸ A reviewer suggested this alternative baseline.

TABLE 4

PARAMETER ESTIMATES OF THE COVARIATES IN THE FULL MIXED MODEL, AN
ALTERNATIVE PERIODIZATION, 1949-89 AND 1990-94

COVARIATES	MODEL 1		MODEL 2	
	Main Effects	Interaction Effects ^a	Main Effects	Interaction Effects ^a
Intercept	2.858***	.537**	2.854***	.529*
Female	-.185***	-.001	-.188***	.000
Age062***	-.013**	.061***	-.013**
Age ² /100	-.071***	.018**	-.068***	.017**
Education:				
Junior high031†	.056**	.035*	.055*
Senior high077***	.086***	.083***	.085***
College136***	.059*	.153***	.054†
Occupation:				
High-rank cadre220***	.059
Low-rank cadre097***	.058*
High-rank professional133***	.033
Low-rank professional045*	.070**
Cadre in public sector141***	.080*
Cadre in economic sector096***	.052†
Professional in public sector082**	.070†
Professional in economic sector049*	.068**
Clerk013	.049	.019	.053†
Service worker032†	-.003	.032†	-.003
Skilled worker008	.003	.008	.002
Work organizations:				
Government agency091***	.053*	.074**	.041
Public organization085***	.051*	.065*	.046
Central government firm136***	.020	.136***	.020
Local government firm084***	.008	.085***	.008
Private/hybrid firms324***	.135***	.323***	.136***
G-covariance structure:				
V _{city}044**	.021**	.044**	.021**
μ _{year}034*033*	...
R-covariance structure:				
σ209***209***	...
ρ	-.00006***	...	-.00009***	...
ε024***025***	...
N	4,730		4,730	
N × T	27,392		27,392	

NOTE.—“Elementary school education or below” is the reference category for education, “unskilled workers” for “occupation,” and “collective firm” for “type of work organization.”

^a The interaction effects indicate changes in the effects of the particular covariates in the 1990-94 period.

† $P < .10$.

* $P < .05$.

** $P < .01$.

*** $P < .001$.

Comparison with the OLS Results and Other Studies

Comparing the parameter estimates based on cross-sectional and panel data (tables 2 and 3), we find that the directions of effects for key covariates are largely consistent in these two sets of analyses. This is comforting evidence that findings based on cross-sectional data analysis in previous studies parallel those based on the panel data analysis. The findings based on the panel data analysis are also similar to other published studies in the urban setting. For instance, although model specifications and data vary across these studies, the estimated returns to party membership and positional power in this study are similar to those reported in Bian and Logan (1996), Xie and Hannum (1996), and Walder (1995*a*). The effects of organizational hierarchy in table 3 are also consistent with the findings on "work-unit characteristic" in Bian and Logan (1996). The estimated returns to education in the reform era (in table 3) appear to be higher than those reported in other studies. This may be due to several reasons: the panel data analysis may be better able to assess changes over time and to untangle the covariation between education, occupation, and work organizations; also, our data captured more recent changes in the course of economic transformation.¹⁹

As indicated in table 2, there are considerable variations in the effects of the covariates across years. These variations may reflect the confounding effects of substantive changes in the reform, temporal changes in response to shifts in state policies, and sample variability. As we see in table 2, choices of different years may lead to different conclusions about changes and stability over time. In contrast, panel data and the mixed model specified here have advantages in addressing the theoretical issues while controlling for random fluctuations and unmeasurable individual attributes over time.

DISCUSSION

Building on theoretical ideas in previous studies, I proposed to conceptualize institutional changes as processes of coevolution between politics and markets and to focus on the interaction between the multiple processes to understand changes in patterns of income determinants in China's economic transformation. My analytical focus on urban China provided a more appropriate setting to assess institutional changes in the transforma-

¹⁹ I do not compare my findings with those reported in the studies of rural China because the institutional arrangements and redistributive channels between rural and urban China are considerably different to allow for a meaningful comparison.

tion of state socialism because the redistributive economy has been more entrenched in urban areas than in rural areas. The use of panel data analyses and the operationalization of key theoretical variables also have improved over the research designs in the previous studies.

It is worth pointing out that the estimated patterns of income determinants for key theoretical variables (returns to positional power and returns to type of organizations) in the prereform era were consistent with the main theoretical arguments of the state socialist stratification order in the literature. This is important evidence that the operationalization of key concepts in this study is consistent with the theoretical specifications, and that the empirical findings (the main effects) establish a baseline for comparing changes in the effects of these covariates across the periods.

The empirical evidence revealed some noticeable changes between the prereform era and the reform era. The strongest evidence consistent with market transition theory is the significant role of private/hybrid firms in the new income regime. These new organizational forms reflect institutional changes that break away from the traditional redistributive economy. Given that this type of work organization is most closely associated with market transactions, I infer that marketization processes have indeed altered channels of resource distribution in urban China. Also, increasing returns to education may be partly attributed to emerging labor markets that better realize values of human capital than before.

On the other hand, the evidence also showed strong institutional persistence. First, regarding returns to political capital and positional power, there is no evidence that returns to party membership or cadre status (in both measures) "decline in significance." That is, we did not observe significant changes in returns to the "redistributors" (high-rank cadres) relative to "producers" (skilled and unskilled workers).²⁰ Second, except for the salient benefits of employees in private/hybrid firms, there were no significant changes in the organizational hierarchy across the two periods. Employees in collective firms, which were closer to market activities than those in the state sector, had not improved their income relative to those in the state sector.

²⁰ A reviewer commented that the findings that, in the reform era, there is no change in returns to the cadre status and that there is a significant increase in returns to education may be interpreted as "the relative decline" of the cadre status. But in the formulation of statistical estimation and tests, the effects of the cadre status should be compared only with the corresponding reference category (unskilled workers). Whether cadres gain (or lose) *disproportionately* with respect to "returns to education" remains to be examined. My exploratory analysis does not show any conclusive evidence on this issue.

To put these findings in a broader context, I note that changes in income determinants were most sensitive to changes in economic activities and captured the most salient aspects of economic transformation in urban China. Furthermore, our data contains information up to 1994, with much more recent evidence than that used in previous studies. In addition, our sampling scheme has a larger proportion of urban residents in large cities, which tend to change more rapidly than other types of cities in China's economic reform. All these factors should help us capture more recent and more dramatic changes due to the expansion of markets and, in this sense, they should favor hypotheses derived from market transition theory.

These findings and my interpretations by no means imply that there were no significant changes in urban China in the last two decades. Quite the opposite. Changes have been both fundamental and widespread. However, these changes were multifaceted, as were the processes that generated them. For instance, increasing returns to education in the reform era clearly reflect the increasing importance of human capital in market transactions. But it is also consistent with state policies in personnel management in the reform era. Similarly, even when we observe similar benefits associated with governmental agencies and positional power of cadres, the sources of these benefits may be partly related to market transactions, as seldom does any work organization or individual in the public sector nowadays solely depend on redistribution from the state. In this sense, although the observed patterns of income determinants may appear to have not changed in significance across the two periods, the processes that generate the apparent continuity may have changed substantively.

These observations raise questions about the usefulness of a focus on discrete institutional forms in understanding institutional changes in China. In the processes of coevolution between markets and the state (and local governments), both are transformed by each other to such an extent that, at an empirical level, conventional concepts may no longer be meaningful to guide our interpretation. Consider the notion of "redistributors." As the state monopoly of resources declines significantly, the political power of the cadres becomes increasingly derived from either the role of "regulators" (as in governmental agencies) or the role of "agents" of the state who manage the production processes (as in state-owned firms). These administrators may still profit from their positional power, which is derived from in part by their association with the political authorities and in part through their involvement in economic transactions in the marketplace. Similarly, although human capital plays an increasingly important role in obtaining economic benefits, the processes of obtaining human capital in the reform era, as Zhou, Moen, and Tuma (1998) shows,

are strongly tilted in favor of the children of those with political or cultural capital rather than ordinary "producers."

These considerations point to a set of challenges in understanding institutional changes in China and other former state socialist societies. First, we need to take a fresh look at both "old" and "new" institutional forms. We need to go beyond labels and ask to what extent the observed economic transactions or "new" institutional phenomena (e.g., business groups, subcontracting) are governed by market or political principles and in what ways they erode, reinforce, or transform markets and politics? In this regard, we need to have a better understanding of how institutional rules are made in the political arena and, in particular, *the specific ways* in which the "old" and "new" interests exert their influence in the rule-making processes. Such analyses would allow us to better understand the sources of advantages and disadvantages among different social groups on a substantive ground.

Second, the persistent and even increasing income inequality revealed in this study is closely related to the critical role of work organizations. Along with the diminishing role of central government in directly managing production, the distribution of economic resources is increasingly determined by one's work organizations. These work organizations are being transformed into new institutional forms whose governance is not necessarily consistent with the principles of either redistribution or markets. Thus it is unlikely that organization-based income inequality can be lessened, let alone eradicated, by the presence of labor markets. Without a good understanding of the new institutional mechanisms of distribution and authority relationships in the workplaces, it is unlikely that we can arrive at a satisfactory answer to sources of income inequality and social stratification in China's transitional economy.

More importantly, the retreat of the redistributive state does not necessarily imply an advance of markets. The emergence of new institutions often takes the form of a recombination of existing routines, authority relationships, and available solutions—such as networks, local corporatism, clientele relationship—within the current institutional arrangements (Bian 1997; Boisot and Child 1996; Lin 1995; Oi 1992; Stark 1996). We need to consider the rise of alternative institutional forms that transcend both redistribution and markets and their implications for the emerging stratification order. Substantive institutional analyses, in my view, require an approach that is historically informed and sensitive to specific institutional contexts, without losing sight of the broad (and competing) causal processes underlying the ongoing economic transformation. Only on this basis can we provide more satisfactory explanations of the observed patterns of income determinants and of institutional changes.

APPENDIX

Technical Report on Model Specification

My main theoretical interest is in examining the effects of the covariates specified in the fixed effects vector, X , in the mixed model (eqq. [1] and [2] in the text)

$$Y = X\beta + Z\gamma + \epsilon, \quad (\text{A1})$$

with

$$\text{var}(Y) = ZGZ' + R, \quad (\text{A2})$$

where Z is the random effects vector, G is the variance/covariance structure associated with Z , and R the variance/covariance structure for time-series correlations among repeated measures within a subject.

Parameter estimates of the fixed effects are sensitive to model specifications in other parts of the model, especially the specification of variance/covariance structures in G and R . In this appendix, I describe and discuss my decision in the model specification of equation (5) in the text. My discussion is concerned mainly with specification of covariates in the X or Z vectors, and the variance/covariance structures of G and R .

Random Effects or Fixed Effects?

The modeling strategy in specifying a mixed model allows one to treat the effects of the covariates as either fixed or random. That is, we need to make a decision about whether a covariate should be put in the X vector or in the Z vector.

Statistically, the choice of treating the effect of a covariate as fixed or random depends on whether one makes inferences conditional on the individual characteristics or makes unconditional inferences on the population characteristics (Hsiao 1986, p. 42; Judge et al. 1985, chap. 13). In general, an effect can be treated as *fixed* if the levels in the sample represent all possible levels of the covariate or at least all levels about which inference is to be made. Conversely, if the levels of the covariate that are used in the study represent only a random sample of a larger set of potential levels, it is appropriate to treat their effects as *random*. Since the main covariates of theoretical interest—age, education, occupation groups, and type of workplaces—are well established in the literature and the main levels in each covariate are measured in the sample as specified in the research design (see the section on variables), I treat these covariates as having fixed effects and include them in the X vector.

I specify the set of indicator variables for respondents' city localities and the selected years in which income data were collected as having

random effects. That is, I assume that the effects of city localities and the selected years are drawn from a larger population of effects associated with city localities and years. Given that both the selected years and cities in the data are only a sample of the populations and that I use years and city localities to control for variations in urban economies over time, this decision is appropriate.

Specification of the R-Covariance Structure

In equation (A2), R refers to the covariance structure associated with the serial correlations among the repeated measures within a subject. Intuitively, we expect the presence of serial correlations among repeated measures within a subject in panel data. Therefore, we need to model the covariance structure of R .

It is common to assume that the effects of the repeated measures decline over time. In this scenario, the conventional autoregressive covariance structure (e.g., AR[1]) is an obvious choice. But, the repeated measures in the data are unequally spaced and vary in the number of years measured depending on the respondents' duration in the labor force. The AR(1) error structure and other time-series covariance structures are inappropriate because they assume equal spacing among repeated measures. I model the serial errors using spatial covariance structures, which are free of the equal-spacing assumption. Moreover, they can be used as a generalization of covariance structures for the one-dimensional time-series error correlations. In particular, the spatial power structure is a direct generalization of the AR(1) structure for unequally spaced repeated measures, and it is intuitively appealing.

I tested the specification of the covariance structures using likelihood ratio tests. I began with a baseline model with independent errors (model 1 in table A1). This model assumes that the errors associated with ϵ_i are i.i.d. with no cross-sectional or serial correlations.

Model 2 introduced the specification of an autoregressive time-series variance/covariance structure (the spatial power specification), which greatly improves the model fit, with two df .

It is likely that there are measurement errors across subjects. One way to model measurement errors is to specify an additional variance term for each subject, which is estimated together with the ϵ_i term in R (Diggle, Liang, and Zeger 1994, p. 87). Model 3 shows that the inclusion of the measurement error term significantly improves the model fit, with 1 df .

Specification of the G-Covariance Structure

In equation (A2), G refers to the covariance structure associated with the covariates in the random-effect vector. In the model specification, I as-

TABLE A1
TESTS OF MODEL SPECIFICATION FOR THE FULL MIXED MODEL

Model Specification ^a	REML Log- likelihood Statistic	χ^2	df
1. Baseline model: $Z = 0$; $R = \sigma I_n$	-22,667.4
Specification of the R structure:			
2. Adding spatial power autoregressive error structure: $R = \sigma^2 \rho^{ 1- 2 }$; $Z = 0$	-12,892.6	19,549.6	2
3. Adding a measurement error parameter: $R = \tau^2 I + \sigma^2 \rho^{ 1- 2 }$; $Z = 0$	-12,451.4	882.4	1
Specification of the Z sector: ^b			
4. Adding only year indicator variables in Z^c : $R = \tau^2 I + \sigma^2 \rho^{ 1- 2 }$; $Z = Z_{\text{year}}$; $G = \sigma_{\text{year}} I$	-11,736.1	1,430.6	10
5. Adding only city indicator variables in Z^c : $R = \tau^2 I + \sigma^2 \rho^{ 1- 2 }$; $Z = Z_{\text{city}}$; $G = \sigma_{\text{city}} I$	-11,701.6	1,499.6	39
6. Adding both city and year indicator variables in Z : $R = \tau^2 I + \sigma^2 \rho^{ 1- 2 }$; $Z = (Z_{\text{year}}, Z_{\text{city}})$; $G = [\sigma_{\text{city}} I, \sigma_{\text{year}} I]$	-10,872.6	1,658.0	10

^a Refer to eqq. (1) and (2) and the appendix for information on model specifications.

^b Degrees of freedom refer to the previous model as the baseline model, except that in both models 4 and 5, degrees of freedom refer to model 3 as the baseline model.

^c Z_{city} refers to the set of indicator variables for city localities; Z_{year} refers to the set of indicator variables for selected years.

sume that G follows a variance-component structure. That is, I specify each of the 20 cities and each of the selected years in the sample to have a distinct variance component to account for heterogeneous urban economic conditions across cities and over the selected years. The variance-component structure also assumes that the variations among these cities or across the selected years are independent of each other.

Models 4 and 5 in table A1 included the random effects of "city localities" and "years" separately with their respective variance-component structures. The likelihood ratio tests show that, in both models, the inclusion of Z and G significantly improves the model fit over the previous model (model 3).

Model 6 included both city localities and years into the random effect vector, and it shows a significant improvement in model fit over model 5. These considerations lead to the full model, model 6, as specified in equation (5) in the text.

REFERENCES

- Becker, Gary S. 1964. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: Columbia University Press.

- Bian, Yanjie. 1994. *Work and Inequality in Urban China*. Albany: State University of New York Press.
- . 1997. "Bringing Strong Ties Back In: Indirect Connection, Bridges, and Job Search in China." *American Sociological Review* 62:266–85.
- Bian, Yanjie, and John W. Logan. 1996. "Market Transition and the Persistence of Power: The Changing Stratification System in Urban China." *American Sociological Review* 61:739–58.
- Boisot, Max, and John Child. 1996. "From Fiefs to Clans and Network Capitalism: Explaining China's Emerging Economic Order." *Administrative Science Quarterly* 41:600–628.
- Campbell, John L., J. Rogers Hollingsworth, and Leon N. Lindberg. 1991. *Governance of the American Economy*. Cambridge: Cambridge University Press.
- Diggle, Peter J., Kung-Yee Liang, and Scott L. Zeger. 1994. *Analysis of Longitudinal Data*. Oxford: Clarendon Press.
- Esping-Andersen, Gøsta. 1990. *The Three Worlds of Welfare Capitalism*. Princeton, N.J.: Princeton University Press.
- Evans, Peter B., Dietrich Rueschemeyer, and Theda Skocpol. 1985. *Bringing the States Back In*. Cambridge: Cambridge University Press.
- Fligstein, Neil, and Iona Mara-Drita. 1996. "How to Make a Market." *American Journal of Sociology* 102:1–33.
- Gerber, Theodore P., and Michael Hout. 1998. "More Shock than Therapy: Market Transition, Employment, and Income in Russia, 1991–1995." *American Journal of Sociology* 104:1–50.
- Granovetter, Mark. 1985. "Economic Action and Social Structure: The Problems of Embeddedness." *American Journal of Sociology* 91:481–510.
- Guthrie, Douglas. 1997. "Between Markets and Politics: Organizational Response to Reform in China." *American Journal of Sociology* 102:1258–304.
- Hamilton, Gary G., and Nicole Woolsey Biggart. 1988. "Market, Culture, and Authority: A Comparative Analysis of Management and Organization in the Far East." *American Journal of Sociology* 94:S52–S94.
- Hodgson, Geoffrey M. 1994. "The Return of Institutional Economics." Pp. 58–76 in *The Handbook of Economic Sociology*, edited by N. J. Smelser and R. Swedberg. Princeton, N.J.: Princeton University Press.
- Hsiao, Cheng. 1986. *Analysis of Panel Data*. New York: Cambridge University Press.
- Judge, George G., W. E. Griffiths, R. Carter Hill, Helmut Lutkepohl, and Tsoung-Chao Lee. 1985. *The Theory and Practice of Econometrics*. New York: Wiley.
- Konrad, G., and Iván Szelényi. 1979. *Intellectuals on the Road to Class Power*. New York: Harcourt Brace Jovanovich.
- Lieberthal, Kenneth G., and David M. Lampton. 1992. *Bureaucracy, Politics, and Decision Making in Post-Mao China*. Berkeley and Los Angeles: University of California Press.
- Lin, Nan. 1995. "Local Market Socialism: Local Corporatism in Action in Rural China." *Theory and Society* 24:301–54.
- Lin, Nan, and Yanjie Bian. 1991. "Getting Ahead in Urban China." *American Journal of Sociology* 97:657–88.
- Littell, Ramon C., George A. Milliken, Walter W. Stroup, and Russell D. Wolfinger. 1996. *SAS System for Mixed Models*. Cary, N.C.: SAS Institute.
- Mincer, Jacob. 1974. *Schooling, Experience and Earnings*. New York: Columbia University Press.
- Naughton, Barry. 1995. *Growing Out of the Plan: Chinese Economic Reform, 1978–1993*. New York: Oxford University Press.
- Nee, Victor. 1989. "A Theory of Market Transition: From Redistribution to Markets in State Socialism." *American Sociological Review* 54:663–81.

- . 1991. "Social Inequalities in Reforming State Socialism: Between Redistribution and Markets in State Socialism." *American Sociological Review* 56:267–82.
- . 1992. "Organizational Dynamics of Market Transition: Hybrid Forms, Property Rights, and Mixed Economy in China." *Administrative Science Quarterly* 37: 1–27.
- . 1996. "The Emergence of a Market Society: Changing Mechanisms of Stratification in China." *American Journal of Sociology* 101:908–49.
- Nee, Victor, and Peng Lian. 1994. "Sleeping with the Enemy: A Dynamic Model of Declining Political Commitment in State Socialism." *Theory and Society* 23:253–96.
- Nee, Victor, and Rebecca Matthews. 1996. "Market Transition and Societal Transformation in Reforming State Socialism." *Annual Review of Sociology* 22:401–35.
- North, Douglass, and Robert Thomas. 1973. *The Rise of the Western World*. Cambridge: Cambridge University Press.
- Oberschall, Anthony. 1996. "The Great Transition: China, Hungary, and Sociology Exit Socialism into the Market." *American Journal of Sociology* 101:1028–41.
- Oi, Jean. 1992. "Fiscal Reform and the Economic Foundations of Local State Corporatism in China." *World Politics* 45:99–126.
- Parish, William L. 1984. "Destratification in China." Pp. 84–120 in *Class and Social Stratification in Post-Revolution China*, edited by J. Watson. New York: Cambridge University Press.
- Parish, William L., and Ethan Michelson. 1996. "Politics and Markets: Dual Transformations." *American Journal of Sociology* 101:1042–59.
- Peng, Yusheng. 1992. "Wage Determination in Rural and Urban Industrial China." *American Sociological Review* 57:198–213.
- Róna-Tas, Ákos. 1994. "'The First Shall Be Last?' Entrepreneurship and Communist Cadres in the Transition from Socialism." *American Journal of Sociology* 100:40–69.
- Shirk, Susan. 1993. *The Political Logic of Economic Reform in China*. Berkeley and Los Angeles: University of California Press.
- SSB (State Statistical Bureau, People's Republic of China). 1990a. *Yearbook of Chinese Cities*. Beijing: China Statistical Publishing House.
- . 1990b. *Historical Statistics of Provinces, Autonomous Regions, and Municipalities, 1949–1989*. Beijing: China Statistical Publishing House.
- Stark, David. 1996. "Recombinant Property in East European Capitalism." *American Journal of Sociology* 101:993–1027.
- Stark, David, and Victor Nee. 1989. "Towards an Institutional Analysis of State Socialism." Pp. 1–31 in *Remaking the Economic Institutions of Socialism*, edited by V. Nee and D. Stark. Stanford, Calif.: Stanford University Press.
- Szelényi, Iván. 1978. "Social Inequalities in State Socialist Redistributive Economies." *International Journal of Comparative Sociology* 16:63–87.
- Szelényi, Iván, and Eric Kostello. 1996. "The Market Transition Debate: Toward a Synthesis?" *American Journal of Sociology* 101:1082–96.
- Wade, Robert. 1990. *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*. Princeton, N.J.: Princeton University Press.
- Walder, Andrew G. 1986. *Communist Neo-Traditionalism: Work and Authority in Chinese Industry*. Berkeley and Los Angeles: University of California Press.
- . 1987. "Wage Reform and the Web of Factory Interests." *China Quarterly* 109: 22–41.
- . 1992. "Property Rights and Stratification in Socialist Redistributive Economies." *American Sociological Review* 57:524–39.
- . 1995a. "Career Mobility and the Communist Political Order." *American Sociological Review* 60:309–28.

- . 1995b. "Local Governments as Industrial Firms." *American Journal of Sociology* 101:263–301.
- . 1996. "Markets and Inequality in Transitional Economies: Toward Testable Theories." *American Journal of Sociology* 101:1060–73.
- Whyte, Martin King, and William L. Parish. 1984. *Urban Life in Contemporary China*. Chicago: University of Chicago Press.
- Xie, Yu, and Emily Hannum. 1996. "Regional Variation in Earnings Inequality in Reform-Era Urban China." *American Journal of Sociology* 101:950–92.
- Zhou, Xueguang, Phyllis Moen, and Nancy Brandon Tuma. 1998. "Educational Stratification in Urban China: 1949–1994." *Sociology of Education* 71:199–222.
- Zhou, Xueguang, and Olga Suhomlinova. 1999. "Redistribution under State Socialism: A USSR and PRC Comparison." *Research in Social Stratification and Mobility*. In press.
- Zhou, Xueguang, Nancy Brandon Tuma, and Phyllis Moen. 1996. "Stratification Dynamics under State Socialism: The Case of Urban China, 1949–1993." *Social Forces* 74:759–96.
- . 1997. "Institutional Change and Job-Shift Patterns in Urban China, 1949–1994." *American Sociological Review* 62:339–65.

Comment: Controversies and Evidence in the Market Transition Debate¹

Yang Cao and Victor G. Nee
Cornell University

Despite repeated attempts to integrate competing perspectives (Szelenyi and Kostello 1996; Nee and Matthews 1996), the ongoing market transition debate has shown no signs of resolution. Instead, the 1996 *AJS* market transition symposium seems to have created more controversy than it settled (Nee 1996; Xie and Hannum 1996; Oberschall 1996; Parish and Michelson 1996; Walder 1996; Fligstein 1996; Szelenyi and Kostello 1996). And subsequent studies continue to reach nearly opposite conclusions (cf. Bian and Logan 1996; Gerber and Hout 1998; with Brainerd 1998; Nee and Cao, in press). "When arguments become polarized, it often signals that divisions are falsely drawn" (Bates 1997). Although originally made in another context, this observation is applicable here. As principals in this lively debate, we believe that clarification and reevaluation are essential for moving toward a reconciliation of competing viewpoints. In this comment we therefore identify the central issues in the controversy and provide an overall assessment of existing empirical evidence.

THEORETICAL ISSUES

At the center of the debate is market transition theory, which attributes the main impetus of change in mechanisms of stratification in postsocialist countries to the emergence of market institutions (Nee 1989). This theory consists of three interrelated propositions that emphasize the causal processes associated with market-based power, incentives, and opportunities, respectively (Nee 1989, pp. 666–67). It predicts that these processes will lead to (1) a decline of the advantage of redistributive power and other forms of political capital *relative* to nonstate economic actors who possess market power;² (2) higher returns to human capital than under a centrally

¹ We thank Mary C. Brinton, Robin Williams, Jr., and Frank W. Young for their comments on an earlier draft. Direct correspondence to Yang Cao, Department of Sociology, 323 Uris Hall, Cornell University, Ithaca, New York 14853.

² "Redistribution" refers to the power of political actors in centrally planned economies to distribute goods and services through administrative fiat and other nonmarket mechanisms.

planned economy; and (3) new opportunities centered on market activities, for example, entrepreneurship. These predictions, however, were soon challenged by many on both theoretical and empirical grounds.

The Declining Significance of Redistributive Power

Although often criticized as an "economistic approach" (Parish and Michelson 1996, p. 1045), market transition theory stresses the change of power structure between the state socialist redistributive elite and economic actors as a key process in shaping the postsocialist stratification order (Nee 1989, p. 666).³ First, to the extent transactions are no longer governed by redistribution but by markets, new opportunity structures emerge *outside* the state socialist redistributive economy in which political actors (i.e., cadres, party members) allocate goods and services. Obviously, this requires the formation of a relatively autonomous market sector, either in the form of secondary economy or through the diversification of formal ownership. Entrepreneurs, managers, and professionals in this market sector are likely to challenge the redistributors' monopoly over elite positions, resulting in a relative decline in the advantage based on political power. Second, if exchanges *between* the state and economic actors are at least partially coordinated by market principles, the latter will be able to retain a greater share of the surplus. In a market economy, producers are not compelled to sell their product or labor at state-mandated prices as previously under a command economy. In the case of rural China, even though the state maintains its dominance in setting the terms of exchange for many agricultural products, peasants are now free to respond to the state's prices by retaining more for consumption, by selling to private merchants, by allocating more time to sideline productions, or by simply producing less.⁴

In the urban economy, exchanges between the state and economic actors differ significantly. Given the indivisibility of the means of production in most urban industrial and service sectors,⁵ urban marketization cannot follow the rural reforms by creating a quasi-commodity market

³ In its original formulation, this thesis partly overlaps with the market opportunity thesis (Nee 1989, pp. 666–67).

⁴ These options may be pursued after fulfilling the mandatory quota, with which most rural households have no problem. A nascent trend in many regions of China is that rural residents with lucrative nonfarm jobs often hire fellow peasants to farm their land, in order to fulfill the mandatory quota.

⁵ The indivisibility of production capital is also among the factors giving rise to path dependent institutional change and local corporatism in rural China (Oi 1990; Nee and Cao, in press).

arrangement between the state and economic actors. Instead, changes in the power structure in the urban economy necessitate the emergence of labor markets, which would allow people to withdraw their labor inputs for employment opportunities elsewhere in case of inferior compensation. Thus, unlike Xie and Hannum (1996), who do not seem to view the emergence of labor markets as an integral part of economic marketization,⁶ we posit that the bargaining power of urban economic actors relative to the redistributive elite is unlikely to be augmented in the absence of labor markets. As a middle-range theory, market transition theory is not limited to rural settings, but specific predictions derived in rural contexts must be qualified before the theory can be tested informatively in urban settings (Cohen 1989).

Indeed, in testing market transition theory's predictions, most empirical research on rural stratification reports confirmatory evidence, while results from urban studies are more mixed (see table 1). The reason, we suspect, is that agricultural decollectivization combined with private entrepreneurship represents a relatively simple form of rural marketization, whereas urban transitions are characterized by much greater path dependence due to both policy reasons and the structural complexity of preexisting urban society. The inertial forces rooted deeply in socialist institutions not only provide the structural basis for the persistence of political advantage under a partial reform (Bian and Logan 1996), but also create conditions more conducive to a variety of strategic adaptations by the state socialist elite (Róna-Tas 1994). Hence, studies pinpointing the importance of specific institutional arrangements or processes contribute especially to our understanding of the socioeconomic transformations in post-socialist countries.

State-Centered Approach

Two lines of state-centered argument have been developed to refute market transition theory's claims regarding the declining significance of redis-

⁶ In their study of income determination in urban China, Xie and Hannum (1996) used economic growth to approximate marketization and found neither lower return to political capital nor higher return to education in more "marketized" cities, i.e., those that experienced faster economic growth. After attributing these findings to the lack of labor markets, Xie and Hannum concluded that their "negative results serve to reject the applicability of Nee's market transition theory, at least to urban China in 1988" (p. 984). Xie and Hannum employed urban survey data collected in 1988 for their study. In our view, the decision to use economic growth to approximate market transition, especially in the beginning stages of urban reform when labor markets and other market institutions in cities had yet to emerge, is highly questionable. Furthermore, if there had been indeed very little marketization in labor allocation, then essentially Xie and Hannum were testing the effects of a process that was yet to take place.

TABLE 1

SUMMARY OF EMPIRICAL STUDIES RELEVANT TO MARKET TRANSITION THEORY

Study	DATA			FINDINGS				
	Nation	Population	Year	Dependent Variable	Political Capital	Human Capital	Entrepreneurship	Private/Hybrid Sector
Szelényi (1988)	Hungary	Rural, whole nation	1982-83	Agricultural goods	Mixed	Yes
Nee (1989)	China	Rural, Fujian Province	1985	Income	Yes	Yes	Yes	...
Nee (1991)	China	Rural, Fujian Province	1985	Income	Yes	...	Yes	...
Peng (1992)	China	Urban/rural, selected areas	1986	Income	...	Yes	...	Yes
Róna-Tas (1994)	Hungary	Urban/rural, whole nation	1989,91	Employment	No	Yes
				Income	No	Yes	Yes	...
Domanski and Heyns (1995)	Poland	Urban/rural, whole nation	1987,91	Income	Yes	Yes	...	Inconclusive
Parish, Zhe, and Li (1995)	China	Rural, eastern two-thirds	1993	Employment	Mixed	Yes
				Income	Yes	Yes	...	Yes
Nee (1996)	China	Rural, whole nation	1989-90	Employment	Mixed	Yes
				Income	Yes	No	Yes	...
Parish and Michelson (1996)	China	Rural, whole nation	1988	Employment	No
				Income	No
Xie and Hannum (1996)	China	Urban, whole nation	1988	Income	No	No
Bian and Logan (1996)	China	Urban, Tianjin	1988, 93	Income	No	Yes	...	Yes
Gerber and Hout (1998)	Russia	Urban/rural, whole nation	1991-95	Income	...	No	Yes	...
Brainerd (1998)	Russia	Urban/rural, whole nation	1991-94	Income	...	Yes
Szelényi (1998)	Hungary	Whole nation	1993	Elite membership	Yes	Yes
Nee and Cao (in press)	China	Rural, whole nation	1998	Employment	Yes	Yes
				Income	Mixed	Mixed	Yes	...
Zhou (in this issue)	China	Urban, selected areas	1995	Income	No	Yes	...	Yes

NOTE.—Yes means that the findings are consistent with market transition theory's prediction; no means not consistent; ellipses indicate either inapplicable or results unavailable.

tributive power and political capital. One, known as the *technocratic-continuity hypothesis*, argues that former cadres can rely on their administrative expertise acquired under state socialism to maintain high socioeconomic status in the postsocialist era. According to Róna-Tas, "There is a common meritocratic-technocratic character of both party and entrepreneurial recruitment that is the main source of continuity" (1994, p. 45). However, as Bian and Logan have pointed out (1996, p. 740), there is no inherent contradiction between this hypothesis and market transition theory. Even though it was cadre status that provided the opportunities to acquire administrative expertise, the reason that such expertise remains consequential is its marketability under the new economic regime. In essence, such expertise is a special form of human capital, and the continuity suggested by the technocratic-continuity hypothesis lies in the apparent similarity between stratification patterns that stem from fundamentally different institutional mechanisms. In this sense, this hypothesis can be, and in fact has been, incorporated into the market transition framework (Nee 1991).

Another state-centered line of argument points to the possibility of power conversion (Staniszki 1991; Róna-Tas 1994). It argues that the political power and capital possessed by cadres can be translated into various forms of economic advantage at the point a market economy is being created. Previous to this, participation in the underground second economy provided rewards that were considerably less lucrative; hence cadres had little incentive to convert their political capital. In Eastern Europe, former cadres' advantages include opportunities to acquire state property through privatization, to establish their own ventures, and to employ political connections to secure profits in business operations in the newly established market economy (Róna-Tas 1994). In the case of China, where the reform remains partial and large-scale privatization has yet to take place, the conversion of political power into economic benefits often takes the form of commodifying redistributive power in the emerging market economy (Nee 1991; Bian and Logan 1996) and of informal privatization relying on network ties between political and economic actors (Nee and Su 1995).⁷ Specifically, the coexistence of two different institutional principles in the Chinese economy allows the political elite to exploit the structural holes, either by commercializing their redistributive power or by brokering between the state and economic actors (Burt 1992).

Despite the apparent contradiction, such conversion of political privilege has in fact been recognized by market transition theory since the very

⁷ Recent anecdotal evidence suggests that cadres in both rural and urban China have begun to utilize their strategic positions to transfer collective properties into their private ownership.

beginning (Nee 1989, p. 668). Note that the power-conversion hypothesis focuses on the microlevel; whether or not this hypothesis implies distinctive stratification patterns on the societal level depends on the *extensiveness* of these mechanisms (Szelényi 1998, pp. 99–121). Because the expanding market economy activates a substantially broader base of economic and social participation that contributes to shaping the postsocialist stratification order, it is likely that even though some of the political elite are able to make timely adaptation, cadres as a social group still experience a relative decline.

To illustrate this point, we construct a hypothetical example. Suppose that (1) under state socialism cadres monopolize the elite positions; (2) numerically, they account for 5% of the total population;⁸ and (3) other things being equal, the odds for them to enter the new elite group—the top 5 percentile in the income distribution—is 2.5 times greater than for others in the transitional period.⁹ Algebraic calculation shows that under ceteris paribus conditions, the new elite group will include only 0.55% former cadres, while the other 4.45% are from other occupational backgrounds.¹⁰ Since 0.55% is only 11% of 5%, this means that the majority of the cadre group (89%) drops out of the economic elite. In other words, cadres' advantage as suggested by the power-conversion hypothesis is theoretically interesting and also likely to manifest itself through statistically significant coefficients in empirical analysis, but its overall importance is questionable.¹¹ This hypothesis would contradict market transition theory

⁸ In Róna-Tas (1994, table 4), former cadres and corporate entrepreneurs account for 4.6% and 5.62% of the total sample, respectively.

⁹ This difference in odds is comparable to former cadres' advantage in entering corporate enterprise, as estimated by Róna-Tas (1994, table 7, models 2–5).

¹⁰ Let P_c and P_o denote the probabilities that cadres and noncadres, respectively, enter the top 5% in the postsocialist era. Similarly, let C denote the proportion of the labor force that is from cadre background and has entered this top 5%, and O denote the proportion of the labor force that is from other backgrounds and in this elite. The four assumptions imply:

$$C + O = 5\%, \quad (1)$$

$$C = 5\% \times P_c, \quad (2)$$

$$O = (1 - 5\%) \times P_o, \quad (3)$$

$$P_c/(1 - P_c) = 2.5 \times P_o/(1 - P_o). \quad (4)$$

Solving for the four unknowns P_c , P_o , C , and O , we have $C = 0.55\%$, and $O = 4.45\%$. For half of the former cadres to maintain their elite positions, their odds would have to be 37 times better than others, corresponding to a coefficient of 3.61 in logistic models. (This can be derived by fixing both C and O at 2.5%, and substituting the odds ratio 2.5 with an unknown in the above system of equations.)

¹¹ For instance, our calculation based on Róna-Tas's results (1994, table 7, model 1) suggests that even with their above-average human capital, only 20% of the former

only insofar as massive power conversion were to take place such that former cadres continued to monopolize the elite positions (Walder 1996, p. 1061)—a scenario which is in our view far from realistic. Given that certain nations and regions do exhibit greater continuity than others in the stratification order, the true challenge is to specify what aspects of institutional change facilitate or impede the strategic adaptation of the political elite (Parish and Michelson 1996; Nee and Cao, *in press*).

Theoretical Synthesis and Model Indeterminacy

In "Economic Transformation and Income Inequality in Urban China: Evidence from Panel Data," his attempt here to synthesize the seemingly divergent views on market transition in China, Xueguang Zhou proposes a "coevolution of politics and markets" model. While Zhou seems to agree with the causal processes suggested by both market transition theory and its critics, he adds greater emphasis to the role of state as a powerful macrolevel actor. He argues that the state's response to the growing market sector creates a mediating effect on changes directly induced by markets; and as a result, his coevolutionary model has significant "model indeterminacy" in empirical predictions. He concludes that it is only through "substantive institutional analyses of how these causal processes interact and exert their impacts in the specific institutional context" that the problem of model indeterminacy can be reduced (p. 1142).

We share with Zhou the same underlying idea that outcomes of social changes are shaped by multiple processes that are often incongruent with each other, but we also agree with his acknowledgment that his study is only a first step toward a synthesis. Instead of embracing competing views in a loosely specified politics-and-markets conceptual framework, a synthesis should, we believe, seek to delineate the scope conditions for specific arguments and assess the relative importance of different processes. Without so doing, predictions are bound to be indeterminate.

Zhou's model implicitly builds on Polanyi's ([1944] 1957) conceptual framework, which links continuous state intervention to market emergence. This framework has in fact been central to market transition theory from the onset. In light of Polanyi's conceptual influence on market transition research (Szelényi and Kostello 1996, pp. 1086–94; Nee and Matthews 1996, p. 407), it is difficult to read Zhou's conception of the coevolution of politics and markets as offering a synthesis that substantively expands

cadres in Hungary were able to become corporate entrepreneurs, and they accounted for roughly 16% of this new elite group. Meanwhile, the tremendous economic mobility experienced by people from various backgrounds in rural China has been extensively documented (e.g., Nee and Liedka 1997).

the causal repertoire already laid out in the literature. Furthermore, Zhou's discussion on state policy stands largely *ad hoc*, and the emphasis on the state as a macrolevel actor thus falls short of a systematic account. For instance, no attempt was made to put the government's new emphasis on educational credentials (p. 1145) into theoretical perspective. Although it is possible that this policy was aimed at retaining qualified personnel so that capable administrators would not be easily lured away by market opportunities, the state might well have done so to promote bureaucratic efficiency, regardless of market growth. Consequently, we are still left wondering in what sense state policies (or market arrangements) should be understood as the results of the interaction between the state and markets. To the extent Zhou's "coevolution of politics and markets" remains a black box, model indeterminacy cannot be reduced. Market transition theory, as Gerber and Hout (1998, p. 10) acknowledge, does not suffer from model indeterminacy.

Market Transition Research Program

The crucial question in the market transition research program remains: In what specific ways do states limit and shape market processes in a manner that provides political actors with significant advantages over economic actors? (See Róna-Tas 1994.) Zhou calls for "substantive institutional analysis," which others have long since begun to undertake. For instance, ethnographic studies on rural China have extensively documented the emergence of a hybrid institutional form—local corporatism (Oi 1990, 1992; Nee 1992; Lin 1995). Building on this, the market transition framework has incorporated the concept of path dependent institutional change and further specifies the systematic linkages between specific institutional arrangements and specific stratification outcomes (Nee and Cao, *in press*). Similarly, urban transition literature has also been enriched by numerous studies on economic reconfiguration, market formation, organizational responses, and so on (Stark 1996; Guthrie 1997; Keister 1998; Matthews 1998; Wank 1999).

Implicit in the market transition debate is a research program aimed at two interrelated objectives: to understand the dynamics of institutional change, and to identify the causal processes through which emerging institutions transform the stratification order. Critics of market transition theory are right in pointing out that the theory, especially in its early formulation (Nee 1989), tends to focus more on the second objective (Walder 1996; Fligstein 1996), although recently the market transition framework has been substantively broadened by its proponents' efforts to endogenize institutional change (Nee and Lian 1994; Nee, *in press*; Nee and Cao, *in press*). In this light, even though considerable controversy remains as to

how to conceptualize institutional dynamics, at least part of the market transition debate is moving toward a constructive resolution.

Meanwhile, critics of market transition theory often overlook the necessity to better specify the causal processes directly affecting the stratification order. This conceptual blind spot is partly evidenced in the fact that most empirical studies on urban stratification rely on indirect measures, for example, time and economic growth, to operationalize marketization (Xie and Hannum 1996; Bian and Logan 1996; Gerber and Hout 1998; Zhou, in this issue). Despite the methodological convenience this strategy offers, it fails to take into account the complexity and variability that characterizes virtually all large-scale societal transformations. In the case of urban China, where market reform remains partial and uneven across regions, a general concept of "marketization" is by no mean adequate in terms of specificity.

Existing market transition studies point to three institutional mechanisms as the immediate causes of changes in the stratification system in postsocialist countries. First, the expanding private/hybrid sector and entrepreneurship create new opportunity structures where economic activities are governed by market principles and are largely beyond the control of the state (Nee 1989). Second, the emergence of labor markets entails a different logic—emphasizing human capital (e.g., expertise and work experience) versus political connections—in matching individuals to positions and this in turn leads to changes in reward distribution (Nee 1996). Third, as economic organizations become more dependent on market exchange for key resources, reward distribution within organizations tends to be modified in favor of human capital and productivity (Cao and Nee 1999). To be sure, these three processes constitute only part of the overall institutional transformation, and their implications on the stratification outcome are likely to be mediated by competing mechanisms such as power conversion and the persistent effects of socialist institutions (Róna-Tas 1994; Bian and Logan 1996). But specific knowledge regarding the growth of the private/hybrid sector, labor markets, and organizational environments provides guidelines for understanding where and to what extent cadre advantage declines relative to returns to human capital and to economic actors. In this light, empirical research will benefit enormously from proper identification and measurement of the key dimensions of the emerging postsocialist institutional order.

EMPIRICAL ISSUES

Table 1 above provides a metaanalysis of existing empirical studies relevant to the market transition debate. The last three columns reveal remarkable empirical consensus over market transition theory's predictions

regarding human capital, entrepreneurship, and private/hybrid sector employment. As expected, the main controversy is centered on the significance of redistributive power and other forms of political capital (see table 1).

Redistributive Power and Political Capital: What Constitutes a Relative Decline?

A crucial theme in the market transition literature is that the declining significance of political advantage must be understood in relation to the upward mobility experienced by other social groups (Nee 1996). As noted earlier, we agree with critics that preexisting institutional arrangements may indeed provide a structural basis for the persistence and conversion of political power (Róna-Tas 1994), but we contend that such mechanisms are unlikely to result in aggregates significant enough to overwhelm the effects of an emerging market economy. For instance, in reforming China, the expanding market sector creates new opportunity structures where entrepreneurs, managers, and professionals can excel economically on the basis of their individual initiatives and human capital. It is likely that the gap between the political elite and the least privileged group may widen during the economic reform. However, to the extent that entrepreneurship, private/hybrid sector employment, and human capital confer greater opportunities for socioeconomic advancement, cadres' monopoly over elite positions can no longer be sustained. Furthermore, once labor markets fully develop in urban China, cadres' advantage over common workers will mainly reflect their superior human capital, rather than redistributive power over resources.

Construed as such, the declining-significance-of-redistributive-power thesis does not necessarily require a diminishing regression coefficient for cadres, relative to a certain noncadre group, over the course of market reform. Most researchers use ordinary or unskilled workers as the reference category for occupational dummy variables. Instead, the reference category should be social groups that represent the rising new market-based classes, that is, entrepreneurs, managers, and professionals. The increasing importance of entrepreneurship, private/hybrid sector employment, and human capital must be taken into account in order to put cadre advantage into perspective. Unfortunately, many empirical researchers have overlooked this subtlety.

Evidence from Urban China

Recent studies on urban transition are riddled with seemingly contradictory evidence. For instance, while Xie and Hannum found no significant increase in the income return to human capital in urban China (1996),

Bian and Logan (1996) and Zhou (in this issue) reported exactly the opposite. A similar disagreement in findings on returns to human capital can be found among researchers looking at Russia (Gerber and Hout 1998; Brainerd 1998). How can such discrepancies be reconciled? And more important, what conclusions can we draw from these results? We suspect that Xie and Hannum's findings may not reflect the consequences of market transition for two reasons. First, the data they employed were collected in 1988, when the Chinese market reform arguably had not made significant headway in most urban areas (Bian and Logan 1996). Second, Xie and Hannum used economic growth to approximate marketization. This dubious operationalization of the key independent variable renders their conclusions even more questionable (Nee and Matthews 1996). Using data from a northern Chinese city, Bian and Logan found that income inequality had been exacerbated in the reform period (1996). On one hand, the economic advantage associated with party membership persisted, and the income gap between cadres and others further widened. On the other hand, returns to human capital, market connections, and private/hybrid sector employment also increased substantially.

We applaud Bian and Logan's research design as well as their effort to integrate network perspectives into market transition research, but we would suggest their interpretation of results be qualified in two ways. First, because their occupational classification scheme does not distinguish redistributive power from managerial control (1996, p. 749, table 1), the estimated cadre advantage is likely to be inflated. In particular, since managers, section chiefs, and supervisors in market-oriented public-sector enterprises were grouped together with political cadres and government officials, this occupational category mainly reflected the difference in positional rank, rather than redistributive power per se. Second, given the increasing returns to human capital, market connections, and private/hybrid sector employment, *persistent cadre advantage alone does not constitute negative evidence for market transition theory*. Market transition theory maintains that economic liberalization frees cadres to pursue power and plenty in the transition economy (Nee 1991). Although cadres may remain advantaged in *absolute terms*, they are losing ground because the emerging market economy enables market-based new classes to rise. So long as significant numbers of nonstate economic actors—that is, entrepreneurs, managers, professionals—outstrip the gains made by cadres (current and former), a *relative decline in the advantages* of the redistributive elite has occurred.

Among existing empirical studies on market transition, Zhou's analysis of changing income distribution in urban China stands out for its rigorous research design and methodological sophistication. Using panel data from 20 cities, Zhou reports findings that, in our assessment, are mostly in line

with market transition theory. Specifically, he finds that income returns to human capital have increased dramatically in the reform era,¹² and although there is evidence for augmented cadre advantage, it is overshadowed by the rapid expansion of opportunity structures centered on private/hybrid sector employment. In addition, professionals have experienced significant gains that allow them to keep up with, if not outpace, those of cadres. With regard to apparent cadre advantage, interpretation of results regarding occupational variables should surely be qualified in the same ways that we view Bian and Logan's findings: managers in market-oriented public-sector enterprises ought not be coded as part of the redistributive elite to the extent that market forces actually shape allocation of resources (Szelényi 1978). Also, while in his descriptive statistics Zhou reports that 4% of his 1993 sample contains private entrepreneurs, this category disappears in his regression analysis.¹³ Overall, even though Zhou's analysis also points to other sources of inequality, the primary cause of changes in the stratification system seems unambiguously tethered to the emergence of a market economy in urban China.

THE RUSSIAN EXPERIENCE

Gerber and Hout (1998) analyze data collected in post-Soviet Russia and claim that 14 of 16 hypotheses derived from market transition theory are not supported. In particular, the income analysis shows no evidence for increasing return to human capital from 1991 to 1995. Some of their hypotheses, especially those regarding gender inequality and differences between public-private sectors, are not truly *derived* from market transition theory.¹⁴ Gerber and Hout implicitly treat economic depression, high-levels of unemployment, and hyperinflation as processes intrinsic to market transition. But these same processes may in fact point to the *failure*

¹² Although Zhou suggests that the increasing returns to education is in part attributable to government policy, his research design fails to substantiate this claim.

¹³ We suspect the omission of entrepreneurs from the regression analysis is due to its high correlation with work organization variables, and this is why coefficients for cadres should be contrasted with those for the private/hybrid sector.

¹⁴ They infer, e.g., from a passage in Nee and Matthews (1996, pp. 428–29), the claim that marketization combined with economic growth will reduce the gender wage gap. But the research Nee and Matthews report on in their review of the market transition literature emphasized the *interaction* between marketization and economic growth as the causal mechanism in reducing gender inequality in China. Specifically, the research showed that only when the demand for nonfarm workers outstrip the supply of male workers will market-driven economic development improve the relative wages of women (Matthews and Nee 1995). In Russia, by contrast, the interaction between marketization and economic depression creates quite different conditions, which appear to have worsened gender inequality in Russia.

to establish a market economy in Russia during the early 1990s. Although many of the formal institutional arrangements of a market economy were set up by reformers in 1992, barter, redistribution, and other forms of nonmarket exchange remain overriding features of the Russian economy. In other words, Russia did not yet have a market economy. While Gerber and Hout suggest that it is Russia's "merchant capitalism" that renders market transition theory inapplicable, their findings are not replicated by Brainerd (1998), who employed similar data and model specifications.¹⁵ Instead, Brainerd finds significant increase in return to human capital, measured by education and occupation, in Russia between 1991 and 1994. With conflicting empirical evidence yet to be reconciled, clear conclusions cannot yet be drawn from the Russian experience. In sum, our discussion of the "negative" findings on political capital reported in table 1 above suggests that much of the controversy over the declining-significance-of-redistributive-power hypothesis stems from misconception rather than actual empirical confirmation that market transition benefits primarily the old state socialist redistributive elite.¹⁶ Clarification and reevaluation of theoretical and empirical issues may contribute toward settlement of differences.

CONCLUDING REMARKS

It is true that the market transition framework tends to focus on the role of market forces in shaping the paths of institutional change and stratification order, while its critics have been more provocative and productive in investigating other simultaneous processes. "When arguments become polarized, it often signals that divisions are falsely drawn" (Bates 1997). However, careful reading of the market transition literature reveals that its conceptual framework incorporates the contending state-centered approach as *contingent hypotheses* that specify the institutional environments in which the old state socialist elite can readily convert political to economic capital during the course of market transition (Nee and Lian 1994; Nee and Cao, in press). Hence it is often extremely hard to draw a clear boundary between market transition theory and the rival arguments. In our view, the difference between market transition theory and its critics

¹⁵ Gerber and Hout's data were collected in February 1991, February 1992, March 1993, July 1994, and January 1996. Brainerd's data were collected in May 1991, April, May, and June 1993, and April, May, and June 1994. With the exception of Gerber and Hout's 1991 data only, all data came from a monthly survey conducted by All-Russian Center for Public Opinion Research, or VTsIOM.

¹⁶ See Nee and Cao (in press) for a reanalysis of Parish and Michelson's (1996) study of rural China.

is mainly a matter of emphasis rather than substantive understanding or position. The real danger, therefore, is not the inability of either side to triumph; rather danger lies in the perpetuation of this falsely drawn division and the persistence of mistaken attempts to use one emphasis to negate others.

REFERENCES

- Bates, Robert. 1997. "Area Studies and the Discipline: A Useful Controversy?" *Political Science and Politics* 30:166-69.
- Bian, Yanjie, and John Logan. 1996. "Market Transition and the Persistence of Power: The Changing Stratification System in Urban China." *American Sociological Review* 61:739-58.
- Brainerd, Elizabeth. 1998. "Winners and Losers in Russia's Economic Transition." *American Economic Review* 88:1094-116.
- Burt, Ronald. 1992. *Structural Holes: The Social Structure of Competition*. Cambridge, Mass.: Harvard University Press.
- Cao, Yang, and Victor Nee. 1999. "Remaking Inequality: Market Dependency and Organizational Adaptation in Urban China." Manuscript. Cornell University, Department of Sociology.
- Cohen, Bernard. 1989. *Developing Sociological Knowledge*. Chicago: Nelson-Hall.
- Domanski, Henrik, and Barbara Heyns. 1995. "Toward a Theory of the Role of State in Market Transition: From Bargaining to Markets in Post-Communism." *European Journal of Sociology* 36:317-51.
- Gerber, Theodore, and Michael Hout. 1998. "More Shock than Therapy: Market Transition, Employment, and Income in Russia, 1991-1995." *American Journal of Sociology* 104:1-50.
- Guthrie, Douglas. 1997. "Between Markets and Politics: Organizational Responses to Reform in China." *American Journal of Sociology* 102:1258-304.
- Keister, Lisa. 1998. "Engineering Growth: Business Growth Structure and Firm Performance in China's Transition Economy." *American Journal of Sociology* 104:404-40.
- Lin, Nan. 1995. "Local Market Socialism: Local Corporatism in Action in Rural China." *Theory and Society* 24:301-54.
- Matthews, Rebecca. 1998. "Where Do Labor Markets Come From?" Ph.D. dissertation. Cornell University, Department of Sociology.
- Matthews, Rebecca, and Victor Nee. 1995. "A Fair Day's Wage: Why Institutions Matter in Gender Inequality." Paper presented at the Ninetieth Annual Meeting of the American Sociological Association, Washington, D.C., August.
- Nee, Victor. 1989. "A Theory of Market Transition: From Redistribution to Markets in State Socialism." *American Sociological Review* 54:663-81.
- . 1991. "Social Inequalities in Reforming State Socialism: Between Redistribution and Markets in China." *American Sociological Review* 56:267-82.
- . 1992. "Organizational Dynamics of Market Transition: Hybrid Forms, Property Rights, and Mixed Economy in China." *Administrative Science Quarterly* 37: 1-27.
- . 1996. "The Emergence of a Market Society: Changing Mechanisms of Stratification in China." *American Journal of Sociology* 101:908-49.
- . In press. "The Role of the State in Making a Market Economy." *Journal of Institutional and Theoretical Economics*.
- Nee, Victor, and Yang Cao. In press. "Path Dependent Societal Transformation: Stratification in Hybrid Mixed Economies." *Theory and Society*.

- Nee, Victor, and Peng Lian. 1994. "Sleeping with the Enemy: A Dynamic Model of Declining Political Commitment in State Socialism." *Theory and Society* 23:253-96.
- Nee, Victor, and Raymond Liedka. 1997. "Markets and Inequality in the Transition from State Socialism." Pp. 202-24 in *Inequality, Democracy, and Economic Development*, edited by M. Midlasky. Cambridge: Cambridge University Press.
- Nee, Victor, and Rebecca Matthews. 1996. "Market Transition and Societal Transformation in Reforming State Socialism." *Annual Review of Sociology* 22:401-35.
- Nee, Victor, and Sijin Su. 1995. "Institutions, Social Ties, and Commitment in China's Corporatist Transformation." Pp. 111-34 in *Reforming Asian Socialism: The Growth of Market Institutions*, edited by John McMillan and Barry Naughton. Ann Arbor: University of Michigan Press.
- Oberschall, Anthony. 1996. "The Great Transition: China, Hungary, and Sociology Exit Socialism into the Market." *American Journal of Sociology* 101:1028-41.
- Oi, Jean. 1990. "The Fate of the Collective after the Commune." Pp. 15-30 in *Chinese Society on the Eve of Tiananmen*, edited by D. Davis and E. Vogel. Cambridge, Mass.: Council on East Asian Studies.
- . 1992. "Fiscal Reform and the Economic Foundations of Local State Corporatism in China." *World Politics* 45:99-126.
- Parish, William, and Ethan Michelson. 1996. "Politics and Markets: Dual Transformations." *American Journal of Sociology* 101:1042-59.
- Parish, William, Xiaoye Zhe, and Fang Li. 1995. "Nonfarm Job and the Marketization of the Chinese Countryside." *China Quarterly* 143:697-730.
- Peng, Yusheng. 1992. "Wage Determination in Rural and Urban China: A Comparison of Public and Private Industrial Sectors." *American Sociological Review* 57: 198-213.
- Polanyi, Karl. (1944) 1957. *The Great Transformation*. Boston: Beacon.
- Róna-Tas, Ákos. 1994. "The First Shall Be Last? Entrepreneurship and Communist Cadres in the Transition from Socialism." *American Journal of Sociology* 100 (1): 40-69.
- Staniszki, Jadwiga. 1991. *The Dynamics of the Breakthrough in Eastern Europe: The Polish Experience*. Berkeley and Los Angeles: University of California Press.
- Stark, David. 1996. "Recombinant Property in East Europe Capitalism." *American Journal of Sociology* 101:993-1027.
- Szelényi, Iván. 1978. "Social Inequalities in State Socialist Redistributive Economies." *International Journal of Comparative Sociology* 19:63-87.
- . 1988. *Socialist Entrepreneurs: Embourgeoisement in Rural Hungary*. Madison: University of Wisconsin Press.
- Szelényi, Iván, and Eric Kostello. 1996. "The Market Transition Debate: Toward a Synthesis?" *American Journal of Sociology* 101:1082-96.
- Szelényi, Szonja. 1998. *Equality by Design: The Grand Experiment in Destratification in Socialist Hungary*. Stanford, Calif.: Stanford University Press.
- Walder, Andrew. 1996. "Markets and Inequality in Transitional Economies: Toward Testable Theories." *American Journal of Sociology* 101:1060-73.
- Wank, David. 1999. *Commodifying Communism: Business, Trust, and Politics in a Chinese City*. Cambridge: Cambridge University Press.
- Xie, Yu, and Emily Hannum. 1996. "Regional Variation in Earning Inequality in Reform-Era Urban China." *American Journal of Sociology* 101:950-92.

REPLY: BEYOND THE DEBATE AND TOWARD SUBSTANTIVE
INSTITUTIONAL ANALYSIS¹

Cao and Nee presented an extended commentary on the recent studies of the transformation of state socialism and a defense of Nee's work. Most of their comments are not directly related to my study. I take this opportunity to address some of the issues raised in their commentary that are related to the themes I pursued in my study.

Logic of Explanation and Model Indeterminacy

Cao and Nee lamented that "the ongoing market transition debate has shown no signs of resolution" and that exchange among scholars "seems to have created more controversy than it settled." I concur with this assessment but disagree with their explanations.

Nee's (1989, 1991) early formulation of market transition theory was appealing to researchers because of its parsimony in theoretical logic and explicit empirical implications. A large number of studies since then show that *some* empirical evidence is consistent with his theoretical predictions. There are also noticeable empirical inconsistencies, which would have provided opportunities for theoretical growth by clarifying/modifying the theoretical logics, specifying scope conditions, and developing alternative theoretical explanations. Unfortunately, Nee's response to his critics, on both theoretical and empirical ground, has added more heat than light.

Theoretically, Nee responds to his critics in two ways. The first is a tendency to claim that the issues that his critics have raised—the continuing role of the state, power conversion, and power persistence—have been "incorporated into" or "recognized by" the original formulation of market transition theory. Second, Nee has tried to incorporate new concepts and mechanisms (e.g., social networks, nonmarket institutions, path dependency, etc.) into his theoretical framework. Unfortunately, as these concepts of mechanisms pile up in Nee's arguments, one increasingly loses sight of how they are interrelated and what their implications are for the internal logic of the theory.

In Nee's original formulation, markets and politics (redistribution) are treated as antithetical—the advance of markets *is assumed to* imply a decline of redistributive power, leading to Nee's empirical predictions for changes in the stratification order. Under the same logic, the effectiveness of market mechanisms may be hindered or distorted by the persistence of politics *because of "partial reform"* (Nee 1991, p. 268). One may disagree

¹ I thank Alan Kerckhoff and Nan Lin for their comments on an earlier version of this note.

with Nee's causal arguments, but this set of arguments is at least *internally* consistent within his assumptions.

But as Nee claims other conceptualizations of mechanisms into his framework, his logic of explanation becomes fuzzy and empirical predictions indeterminate. For instance, if power conversion and power persistence emphasized in Róna-Tas (1994) and Bian and Logan (1996) are part of market transition theory, as Cao and Nee suggest above, then Nee implicitly accepts the assumption that politics may be governed by alternative processes, which operate along with (rather than as the antithesis of) markets. This conceptualization opens doors to a wide range of logical possibilities: since political power may be converted to or persist in generating economic benefits *in the marketplace*, it becomes logically possible that those in political positions benefit disproportionately from market transactions (as Róna-Tas and Bian and Logan demonstrated). Thus, the expansion of markets is no longer a sufficient condition for the decline of redistributive power. That is, the decline of political capital/status is no longer the logical consequence of an increase in market mechanisms.

As another example, consider the role of social networks and local corporatism as alternative mechanisms of organizing economic activities (Boisot and Child 1996; Lin 1995; Oi 1992; Walder 1995b). If we treat social networks and local corporatism as *alternative, independent processes* that shape economic transactions, as Nee (1996) appeared to agree, then the implications of these processes on the stratification order inevitably transcend the redistributor-direct producer dichotomy in market transition theory. It is logically possible that those occupying advantageous structural/network positions (rather than affinity with markets) gain disproportionately from economic transactions, leading to empirical implications that are qualitatively different from those derived from the market transition framework.

Recent studies have advanced our knowledge of the field by developing alternative conceptualizations of the change processes—local corporatism, network capitalism, and politics of workplaces—that cannot be simply reduced into Nee's market transition logic. To the extent that these are conceptualized as alternative or competing processes to market mechanisms, they lead to different logics of explanation and empirical predictions. Similarly, my proposed model of coevolution between markets and politics, admittedly underspecified in the article,² can hardly be claimed as part of market transition logic: first, the coevolutionary model treats market mechanisms as but one of the competing processes in institutional changes. In this conceptualization, politics, of which the state is a central

² For substantive implications of coevolutionary processes in biology, where this concept originated, see Durham (1991) and Kaufman (1993).

component, does not merely exert a mediating or second-order effect in the marketization process; rather, it plays an active role in shaping emerging economic institutions. Second, a core idea of the coevolution concept is that, because both markets and politics coevolve in response to each other, the outcome depends on the specific and evolving conditions in the institutional context. In this light, theorizing based on one or the another process alone does not yield any predictive power on the outcome of the coevolutionary process. Thus the coevolutionary model is at odds with market transition theory in both empirical implications and research agendas.

It seems to me that Nee can either acknowledge these alternative, competing processes as an integral part of his theoretical framework and, in so doing, reformulate his logic of explanation and empirical predictions, *or* he can explicate the internal logic and the scope conditions of his theory and be prepared to accept that observed empirical inconsistencies are challenges to, among other things, his theoretic logic of explanation. Unfortunately, in my view, Nee expands the "causal repertoire" of his theory at the expense of conceptual clarity, logical consistency, and empirical falsifiability.

Empirical Evidence

Empirically, Nee's response to his critics (table 1 above) is to claim that empirical evidence is largely consistent with market transition theory. But Nee's selective attention to empirical evidence casts doubts on this claim. Let me illustrate this point by reinterpreting Cao and Nee's interpretation of the findings in my study.

First, the interpretation of returns to human capital. Nee sees higher returns to human capital (education) in China's reform era as evidence in support of market transition theory. But one needs to remember that similar empirical predictions have already existed in theories of industrialization and of modernization in the sociological literature. There are also competing theoretical arguments that human capital is rewarded under state socialism (see Walder [1995*a*], p. 311 for a review). Indeed, as Zhou and Suhomlinova (1999) show, returns to education in the Soviet Union were much higher than those in urban China were on the eve of the great transformations of redistribution in these two societies. In fact, returns to college education before the collapse of the Soviet Union in 1991 were even higher than those in urban China in the reform era, *ceteris paribus* (cf. my table 2 above with table 2 in Zhou and Suhomlinova [1999]). Thus the evidence of higher returns to human capital cannot be construed as caused by market mechanisms without a substantive understanding of how human capital is allocated in the specific institutional context.

Moreover, as I noted in my article, the emphasis on human capital was

part of state wage policies since the early 1980s, before substantive market reforms took place in urban China. Corroborative evidence shows that, in the period of my study, those with high human capital were more likely to stay in or shift to the state sector relative to the nonstate sector (Zhou, Tuma, and Moen 1997). On this basis, I believe my interpretation (in the text) that increasing returns to human capital (education) in China's reform era result from the compound effects of political processes and marketization is more appropriate than Cao and Nee's view that it is solely attributable to market mechanisms. When multiple processes are observed to generate the same empirical outcome, the empirical evidence cannot be used as a critical test of competing theoretical arguments.

Second, I am surprised that Cao and Nee failed to see the salient evidence reported in my study that is inconsistent with the empirical predictions of market transition theory—the noticeably lower income for employees in collective firms than for those in the state sector. Collective firms are closer to market activities relative to work organizations in the state sector because their supply and output have never been an integral part of the command economy in urban China. Nee (1992, p. 7) also recognizes this: "Because collective enterprises fall outside the central plan, which encompasses state-owned enterprises, their growth became increasingly dependent on markets. . . . Collective enterprises quickly oriented their production to the marketplace and contributed to the rapid expansion of markets in China." According to Nee's theoretical logic, we would expect to find higher returns to this group of "direct producers" (or, in Cao and Nee above, "economic actors") who actively engage in and benefit from market activities. The finding of persistent low income for employees in collective firms is as salient and important a piece of empirical evidence as is the high income for the private entrepreneur group, in the empirical test of market transition theory.

Toward Substantive Institutional Analysis

When a theoretical debate generates more controversies than intellectual growth, it often signals that the conceptual issues and theoretical logics are poorly defined and they are not widely shared among scholars. Another contributing factor is that the concepts and operationalization employed in empirical studies may no longer reflect the changing world. My observations of the institutional changes in China make me skeptical of the usefulness of continuing this debate based on those concepts and categories that are becoming increasingly fuzzy and irrelevant in the ongoing processes of change.

Consider the notion of "redistributor." In the ideal-type of state socialist redistribution in which all resources are allocated through the state bu-

reaucracy, it is sensible to define those with bureaucratic power (cadres) as redistributors. In the transformation processes, however, the empirical contents of this concept become highly differentiated. For instance, a manager in the state firm today acts both as the agent of the state and as a manager of the enterprise. It is both conceptually and empirically uncertain to what extent and in what ways his or her observed income is derived from the affiliation with the political authorities or from the managerial role of "director producer." Similarly, the category of "professionals" does not help much, either. In some economic sectors, professionals and their work are closely linked to market mechanisms; in other sectors (state firms, research institutions), professional work and rewards are still heavily influenced by state policies and regulations. Nee (1996) advocates treating managers and professionals as "direct producers." But the conceptual and empirical ambiguities associated with these categories cannot be swept away by a simple reclassification, and uncritical use of these concepts and categories in empirical studies adds more confusion than clarification of the theoretical issues.

Large-scale institutional changes involve multiple causal mechanisms, which require close observations and careful detective work to identify, understand, and untangle in the actual processes of change. For instance, does the presence of a formal contractual relationship between two firms signal that the economic transaction is governed by market mechanisms on a rational-legal basis? Or is it merely a symbolic compliance to state regulations, and the actual processes are governed by social relationships (e.g., social networks), or political mechanisms (brokered by the government)? Or is it a mixture of several processes? Along with large-scale institutional changes under way, the processes that underlie these changes may evolve and take different forms as well. The bottom line is that, without careful and systematic examinations of these processes, we cannot even begin to describe these phenomena accurately, let alone to develop sound theoretical explanations about their roles in institutional changes.

In my article, I advocate moving away from the current debate and toward substantive institutional analysis of the processes of transformation. Institutional changes in the transformation of state socialism demand theoretical explanations. But theorizing should be built on, not substitute for, the knowledge of the specific context and the processes in which these changes take place. On this account, I find those substantive institutional analyses in the literature (e.g., Lee 1995; Lin 1995; Lin and Ye 1998; Oi 1999; Stark, 1996; among others) especially informative, because they direct our attention to seek answers in the actual processes of institutional changes, rather than confine our imagination within the terms of the current debate.

XUEGUANG ZHOU

Duke University

REFERENCES

- Bian, Yanjie, and John Logan. 1996. "Market Transition and the Persistence of Power." *American Sociological Review* 61:739–58.
- Boisot, Max, and John Child. 1996. "From Fiefs to Clans and Network Capitalism: Explaining China's Emerging Economic Order." *Administrative Science Quarterly* 41:600–628.
- Durham, William H. 1991. *Coevolution: Genes, Culture, and Human Diversity*. Stanford, Calif.: Stanford University Press.
- Kauffman, Stuart A. 1993. *The Origins of Order: Self-Organization and Selection in Evolution*. New York: Oxford University Press.
- Lee, Ching Kwan. 1995. "Engendering the World of Labor." *American Sociological Review* 60:378–97.
- Lin, Nan. 1995. "Local Market Socialism: Local Corporatism in Action in Rural China." *Theory and Society* 24:301–54.
- Lin, Nan, and Xiaolan Ye. 1998. "Chinese Rural Enterprises in Transformation: The End of the Beginning." *Issues and Studies* 34:1–28.
- Nee, Victor. 1989. "A Theory of Market Transition: From Redistribution to Markets in State Socialism." *American Sociological Review* 54:663–81.
- . 1991. "Social Inequalities in Reforming State Socialism: Between Redistribution and Markets in State Socialism." *American Sociological Review* 56:267–82.
- . 1992. "Organizational Dynamics of Market Transition: Hybrid Forms, Property Rights, and Mixed Economy in China." *Administrative Science Quarterly* 37:1–27.
- . 1996. "The Emergence of a Market Society: Changing Mechanisms of Stratification in China." *American Journal of Sociology* 101:908–49.
- Oi, Jean C. 1992. "Fiscal Reform and the Economic Foundations of Local State Corporatism in China." *World Politics* 45:99–126.
- . 1999. *Rural China Takes Off*. Berkeley: University of California Press.
- Róna-Tas, Akos. 1994. "The First Shall Be Last? Entrepreneurship and Communist Cadres in the Transition from Socialism." *American Journal of Sociology* 100:40–69.
- Stark, David. 1996. "Recombinant Property in East European Capitalism." *American Journal of Sociology* 101:993–1027.
- Walder, Andrew G. 1995a. "Career Mobility and the Communist Political Order." *American Sociological Review* 60:309–28.
- . 1995b. "Local Governments as Industrial Firms." *American Journal of Sociology* 101:263–301.
- Zhou, Xueguang, Nancy Brandon Tuma, and Phyllis Moen. 1997. "Institutional Change and Job-Shift Patterns in Urban China, 1949–1994." *American Sociological Review* 62:339–65.
- Zhou, Xueguang, and Olga Suhomlinova. 1999. "Redistribution under State Socialism: A USSR and PRC Comparison." *Research in Social Stratification and Mobility*, in press.

Book Reviews

ReOrient: Global Economy in the Asian Age. By Andre Gunder Frank. Berkeley and Los Angeles: University of California Press, 1998. Pp. xxix+416.

Christopher Chase-Dunn
Johns Hopkins University

Andre Gunder Frank's *ReOrient* has caused great waves of anxiety among social scientists because of his claim that this new perspective on the rise of the West invalidates all our theories of development. The great contribution of the book, and its predecessor, *The World System: 500 or 5,000 Years?* (with Barry Gills [Routledge, 1993]) is that it forces us to focus on the significant economic and political/military continuities that have characterized the rise and fall of powerful regions for millennia. Though Frank somewhat overstates the size and degree of integration of the world system that emerged in Afroeurasia, he is undoubtedly correct that interactional linkages importantly conditioned the processes of development in the West Asian-Mediterranean and East Asian regions much earlier than most historians and civilizationists have supposed. Frank is also right to point out that Europe was long linked to this emerging system as a peripheral area, and that China was probably the most powerful and dynamic core region in the system until the late 18th century.

In making these arguments, Frank reiterates the importance of processes at the level of the whole system, and he uses data on the sizes and growth of cities as well as long-distance trade to sustain his claims. He further contends that the rise of European hegemony was the sudden and conjunctural outcome of a developmental crisis in China, which, following Mark Elvin, is characterized as a "high level equilibrium trap." The Chinese economy had expanded demographically such that labor was so cheap that there was no incentive to pursue labor-saving techniques. This occurred after a period in which European states had used the riches extracted from their colonies in the Americas to buy into the Asian economic centrality, and so Europeans were able to take advantage of the Chinese crisis to industrialize and become hegemonic in the world system. Frank contends that European hegemony was fragile from the start and will be short-lived with a predicted new rise of Chinese predominance in the near future.

Frank argues that there were no important differences in developmental logic between Europe and the other competing regions. He explicitly contends that there was no "transition from feudalism to capitalism" in Europe that was the basis of Europe's rise to hegemony. He claims

Permission to reprint a book review printed in this section may be obtained only from the author.

that theories of transitions between modes of production are part of the Eurocentric ideological baggage that has distorted historical social science and that must now be discarded. He also throws out Weberian sociology's allegations about Eastern/Western differences in political organization as so much Eurocentric trash.

It is important to pick carefully through the rubble and to examine the dead bodies of our social science ancestors that litter the ground after Frank's large-caliber massacre. It is also important to ascertain what parts of his argument contain valuable insights from those parts that are mainly rhetorical overkill. We can admit that non-European regions have not received their fair share of attention from social scientists and that some of the allegedly superior attributes of European societies that led to hegemony were ideological justifications for the exercise of domination. But these points do not add up to the conclusion that all of our theories are invalid. Rather, we need to revisit basic theoretical assumptions in the light that is shed by the kind of large-scale systemic analysis and comparisons undertaken by Frank.

In my view, Frank is right that there are many more continuities and systemic links than most historical social scientists have imagined. Evidence on the rise and fall of cities and changes in the territorial sizes of empires indicates that the East Asian and West Asian/Mediterranean regions were experiencing synchronous processes of change after B.C.E. 500. This is most likely a consequence of long-distance trade linkages and military interaction with the state-formation processes of Central Asian steppe nomads (e.g., the Mongol Empire). Frank's emphasis on the rise of a Afroeurasia-wide world system is a valuable hypothesis that will require careful research on interaction networks and the possible causes of synchronicity. Frank's contention that China was the most powerful and dynamic center until the late 18th century may also be true, though there were two other powerful core regions in the system (West Asia/Mediterranean and South Asia), and so it is best understood as a multicore system. Frank's claim that Europe was long a peripheral region to the West Asian/Mediterranean core is also right. The western Roman empire was a temporary exception. But Frank's claim that the European rise to hegemony was a short-term and conjunctural outcome that had no implications for developmental logic is most probably mistaken. European cities had been growing at extraordinary rates since the 14th century, and the emergence within Europe of an interstate system in which the hegemonic core powers were capitalist states instead of tributary empires was most probably due to an important difference in strength of capitalist accumulation in Europe compared with the older core regions in South and East Asia. It was the rise of Europe and the development of capitalist industrialization that created a truly global system with a single integrated group of core states. This is structurally quite a different matter from the multicore Afroeurasian system studied by Frank. *ReOrient* is a stimulating and thoughtful book that should be read by all serious students of the modern world system.

Nationalism, Liberalism, and Progress. Vol. 1, *The Rise and Decline of Nationalism*. By Ernest B. Haas. Ithaca, N.Y.: Cornell University Press, 1997. Pp. ix+362. \$39.95.

John A. Armstrong
University of Wisconsin, Madison

Ernest Haas, a political scientist, presents a broad study of the role of nationalism in societal modernization. This first volume of *Nationalism, Liberalism, and Progress* deals with the five largest industrialized countries: Britain, France, the United States, Germany, and Japan. Haas sets out to correct "the dark view of nationalism as historically simplistic and morally misleading"; it is, rather, "a human invention designed, in a sense, to make life better for collectivities suffering the pangs of modernization" (pp. vii–viii). Haas cites Max Weber (p. 2) as specifying wealth, health, and peace as material improvements people want in addition to "emotional and spiritual happiness," but Haas concentrates on material factors. In this respect Haas is an avowed child of the Enlightenment, not the pessimist Weber often appeared to be. For example, Haas argues that initially "nations result from the appropriate combination of rewards and punishments handed out by rulers, and that people, when given the chance, respond instrumentally to these inducements" (p. 2). Consequently, he rejects the functionalist reasoning that colored the massive interpretation of European states by Gabriel Almond, Stein Rokkan, and Charles Tilly. Haas also starts his investigation at a chronologically later stage than did these authors, contending (I think exaggeratedly) that the earlier territorial state was populated by "impassive and inert" subjects (p. 18). Change, after the late 18th century, was accompanied by material progress, but more crucially by the *idea* of progress through consensual rationalization, that is, policy based on "scientific" reasoning. For relatively brief periods, syncretist elites, even such extremists as the Nazis, could promote consensual rationalization by eliminating barriers like class; ultimately, though, only *liberal* nationalism could promote exchange reciprocity needed to attain the *formal rationality* postulated by Weber.

To systematize his findings, Haas presents 16 indicators, such as predictable rules of succession and popular acceptance of tax payments, military conscription, and official language. Consensus is weighed, for each indicator, on a scale of zero, 0.5, or one, and their sum (adjusted to a percentage of the total) provides an overall indicator of consensual rationalization for a given year. Summations for seven such years (differing for each society) are tabulated in each national chapter. Taken together the summations suggest a learning curve for each society. Clearly, the results reflect the author's personal judgments. These are backed by his great experience and linguistic ability (except perhaps for Japan) as well as quantitative data opinion surveys and longer time series on state budgets. I should have preferred more attention to specific learning mecha-

nisms. When, as in the case of post-1880 France, Haas does analyze the roles of institutions like the École Normale in training and supervising a new "republican" *lycée* teaching corps for inculcating the French language and the values of secular nationalism, his conclusions are more convincing (p. 188).

Several of Haas's conclusions contradict conventional wisdom, at least that current in the United States. At an early point he presents a typology of (A) countries, France and Britain, where the state was well established before nationalism, and (C) countries, United States and Germany, where a nationalist elite produced a state by secession or consolidation. Yet he concludes that a more significant dichotomization characterizes countries where the nation-state has produced a satisfied population, namely the continental countries (and perhaps Japan, although there feelings of insecurity among elites and masses remain apparent) as compared to the English-speaking countries. Haas emphasizes that progress toward consensual rationalization on the Continent was far from smooth or rapid. During the post-World War II decades, however, France and Germany have attained a higher degree of internal reciprocity and now recognize that further progress depends on replacing nationalism by a large measure of allegiance to international institutions, specifically the European Community. In contrast, the Vietnam experience, racial antagonisms, and revived religious dissent from national policies have to some extent "derationalized" U.S. policy making since 1970. Haas sees these difficulties as closely linked to reliance on lawyers as policy makers rather than on "knowledge specialists" (as neglected examples of the latter I suggest Gunnar Myrdal and Hans J. Morgenthau). In Britain, Haas identifies immigration controversies, lagging economic growth, and abandonment of inculcating patriotism in both elite secondary schools and comprehensive schools.

One further consideration, that Haas mentions too briefly: the invidious social Darwinism strong in both English-speaking countries from the late 1890s until at least the late 1920s. These racist and chauvinist doctrines were, to be sure, abandoned in later decades. But they were never confronted in the decisive way in which, after the traumas of defeat, French and German elites were compelled to reject their more extreme xenophobias. By such confrontations, elites, and to a considerable extent popular elements, learned also to renounce bitter divisions over religion—Catholic versus Protestant in Germany, clerical versus Jacobin nationalists in France—in order to avert crises by consensual policy making.

Overall, Ernst Haas has made a remarkable contribution to our understanding of the complex relationships between modernity and nationalism. In his eagerly awaited second volume, he will probe Third World nationalisms in the modernity process.

Fantasy City: Pleasure and Profit in the Postmodern Metropolis. By John Hannigan. London: Routledge, 1998. Pp. xvi+239.

David Grazian
University of Chicago

The central argument in John Hannigan's *Fantasy City* repeats the familiar critique of the recent development of Disneyesque entertainment districts in postindustrial cities around the globe. As downtown complexes and waterfront projects proliferate according to the theme-park model suggested by Boston's Faneuil Hall and New York City's South Street Seaport, they grow increasingly similar by providing the same menu of corporate-owned coffee franchises, mall-savvy haberdasheries, IMAX theaters, themed restaurants, book superstores, and entertainment company outlets. Specifically, Hannigan asserts that, in spite of their popularity among consumers, these spaces of commodified pleasure may prove costly to inner-city populations, resident tax bases, and the overall quality of life among locals and tourists alike. While his narrative offers little more than a derivative compendium of facts, figures, and fun that relies heavily on journalistic accounts, collegiate instructors will find *Fantasy City* useful classroom fodder for its accessibility to undergraduate students.

In terms of production, Hannigan argues that several contemporary trends—notably the increased dominance of rational techniques of production and the rise of synergy as a key logic in the entertainment and development industries—are responsible for the rise of urban entertainment districts. He notes that many rely on the synergistic effects of multiple-use facilities and licensed merchandizing to incite visitors of such projects to buy tickets for a mainstream movie, ride its namesake roller coaster, play accompanying virtual-reality games, and purchase action figures and tacky T-shirts to remind them of all the fun. On the consumption end, he asserts that the seduction of “technologies of simulation” and the rising value of cultural capital based on “experiences” rather than knowledge appeal to postmodern consumers. Furthermore, Hannigan attributes the popularity of fantasy city to its ability to reconcile the contradictory desires among middle-class consumers for “adventurous” yet completely risk-free experiences. Urban malls and themed restaurants succeed because they reassure as they dazzle by combining normally pedestrian activities such as shopping and dining with hyperreal spectacles and the highly dramatized worlds of infotainment and make-believe without the inconveniences of urban blight, the presence of “undesirables,” and the unpredictability of everyday life.

The strength of Hannigan's book rests in its ability to draw on a wealth of specific cases ranging from the notorious to their less obvious derivations. Examples of the latter include the development of a Chuck Norris–endorsed gaming operation in Moscow's Red Square, the global proliferation of Hard Rock Cafe franchises in Kuala Lumpur and Jakarta, and

the recent "shameless" practice of renaming local sports venues after corporate sponsors, including New Jersey's Continental Airlines Arena and Phoenix's Bank One Ballpark. Another strength of his analysis lies in his emphasis on the local variations, which inevitably emerge as themed entertainment spreads to the Asia-Pacific Rim and other foreign ports of call. Likewise, his account of the kinds of public-private partnerships, which regularly fail due to a mismatch of interests among local boosters and other urban elites, suggests an ecological model of local decision making that improves upon perspectives employed by less flexible critics of urban governance. Finally, Hannigan deftly explores the development of fantasy city as a historical process with antecedents in the popular amusements of the 1920s, rather than a consequence of some postmodern zeitgeist. His discussion of their reliance on exclusionary admittance practices and elements of "staged authenticity" seems especially relevant to an analysis of the present.

Still, *Fantasy City* suffers from a number of common problems indicative of much of the literature on this well-seasoned topic. First, Hannigan relies far too heavily on tenuous and essentialist definitions of problematic terms such as "postmodernism" and "authenticity" without offering more nuanced readings of the field based on his own observations. In fact, his very conception of the urban entertainment district itself suggests a tautology insofar as it relies on the same criteria he uses to explain the causes of its emergence in recent years. Also, *Fantasy City* all too often leaves central political issues demanded by his project unexplored: for example, the book includes a chapter devoted entirely to the proliferation of gaming operations intended to spawn urban development, yet fails to address the implications of allowing gambling ventures, which prey on the poor, to grow adjacent to low-income metropolitan regions with the help of government-backed financial assistance.

Finally, like many studies that purport to explain how so-called postmodern consumers interpret the cultural world around them, *Fantasy City* suffers from its lack of ethnographic data on the reception of cultural forms among actual audiences. Certainly, Hannigan moves away from the annoying abstraction and navel gazing of a Jean Baudrillard or Fredric Jameson by orienting his discussion around a more grounded critique of the political economy of the cultural landscape. Still, his lack of attentiveness to empirical qualitative research on the processes of meaning-making employed by the living consumers of fantasy city is disappointing. Indeed, the emergent postmodern culture of Disney and Hollywood may offer highly sanitized experiences, but that hardly means our studies of them should be equally sterile.

Globalisation, Human Rights, and Labour Law in Pacific Asia. By Anthony Woodiwiss. Cambridge: Cambridge University Press, 1998. Pp. xii+316. \$64.95 (cloth); \$24.95 (paper).

Pheng Cheah

University of California at Berkeley

The debate over the normative status of human rights in a post-Cold War global era has generally remained within the parameters of a tired quarrel between universalism and cultural relativism. In this book, Anthony Woodiwiss attempts to articulate an alterantive account of human rights, which, while it does not take the civil and political liberties championed by Western liberalism as its central paradigm, also does not reject the idea of the universality of human rights in favor of cultural relativism. Woodiwiss's rather Hegelian argument is that, in order to be effective, any human rights regime needs to be embedded in the fabric of governance of a given social formation since it is this social embeddedness that makes the regime enforceable. Hence, instead of dismissing the forms of governance in rapidly industrializing Pacific Asia as authoritarian regimes that are inimical to human rights, he suggests that they ought to be regarded in ideal-typical terms as a new form of sovereignty called "patriarchalism"—a form of rule that is exercised in the interests of all members of society, which is viewed in quasi-familial terms. Woodiwiss argues that although patriarchalism is typologically distinct from Western liberal society because it is characterized by quasi-familial hierarchical social relations, it is a form of "enforceable benevolence" that can enhance respect for human rights provided that it is accompanied by and hybridized with democratic structures and the rule of law. Indeed, he makes the stronger claim that authentic patriarchalist regimes may be more effective for the institutionalization of human rights in Pacific Asia than regimes based on individualistic or social-democratic checks on inequality.

This attempt to construct an alternative paradigm for human rights "that combines elements of the socialist, patriarchalist and liberal traditions" (p. 243) is instructive because it refuses to pay lip service to "Confucian values" or "Asian values" that are now a compulsory ideology of most East and Southeast Asian states. Although Woodiwiss draws on the historical rise of Pacific capitalism—in particular, the model of industrial justice found in post-World War II Japan—he pointedly distinguishes his position from spurious culturalist arguments that posit a genetic connection between East and Southeast Asian capitalism and Confucian values (see pp. 152–56). Instead, he takes as his central focus the tension between capital and labor in general. He argues that in industrializing Pacific Asia in a global era, the determinate sociological form generated by and conditioning the struggle between capital and labor is such that labor rights are most effectively enforced against transnational and local capital through judicial recognition of the state's obligations to its citi-

zens. The concept of lifetime employment in the Japanese experience is one instance of this enforceable benevolence.

The second half of the book analyzes labor rights in the Philippines, Hong Kong, Malaysia, and Singapore in terms of how far they measure up to the ideal-type. These case studies variously indicate failure to achieve the ideal of enforceable benevolence, and the four countries are given more specific typological labels: mendicant patriarchalism (the Philippines); patriarchalist individualism (Hong Kong); authoritarian patriarchalism (Malaysia); and active patriarchalism (Singapore). However, Woodiwiss's final evaluation of Singapore is generous: he suggests that given the city-state's provisions for the social welfare of its citizens, it can displace Japan as an achieved instance of enforceable benevolence provided that guarantees of judicial independence and freedom of labor organization are established.

But such a conclusion should not be surprising. For although Woodiwiss reiterates that "capitalism as a form of domestic, international and transnational economic organisation *is* intrinsically subversive of respect for human rights" (p. 247), he also believes that it introduces structures (such as the organization of labor) that can offset some of the human rights violations it brings about. Indeed, the historical basis for the ideal-type of enforceable benevolence is the rapid development of Pacific Asia as a result of its integration into the circuit of global capitalism. Since Singapore is arguably the most economically advanced and socially well-regulated Asian state in the age of disorganized global capitalism, it would seem logical for Woodiwiss to regard it as being "structurally closest to constituting a new human rights regime" (p. 246).

One suspects that some of the conclusions of the book may have to be revised in light of the social upheaval caused by the Asian financial crisis. Moreover, even if one accepts that many of the conditions for safeguarding labor rights are already in place in Singapore, the protection of labor rights within individual nation-states is not identical to a new regime for universally enforceable human rights. In fact, Woodiwiss's focus is almost exclusively national. He focuses on the protection of the rights of workers by their states. He does not discuss one crucial area of worker's rights that has expanded as a result of the relentless search for low-labor costs in Pacific capitalism: the rights of legal and clandestine migrant workers. Three of the regimes he studies (Singapore, Malaysia, and Hong Kong) thrive by exploiting migrant workers to perform the 3D jobs (dirty, dangerous, and demanding) that their domestic workforce has shunned. (Foreign maids in Singapore are required to undergo pregnancy tests once every six months and are immediately deported if found to be pregnant). From a structural point of view, the rights of these foreign workers cannot be protected by the bargaining between capital, state, and labor Woodiwiss discusses because they do not belong to these patriarchalist families. The home countries of migrant workers are also generally reluctant to insist on their protection for fear of losing their share of the labor market. Thus, whatever gains are made by workers who are

nationals within various patriarchalist regimes, they are offset by the inhumane treatment of foreign workers. For the model of enforceable benevolence to develop into a genuine regime of universal human rights, it needs to come to terms with the competitive nature of a global economy that pits the rights of the workers of each Asian nation against the other.

Los Angeles: Globalization, Urbanization and Social Struggles. By Roger Keil. New York: John Wiley, 1999. Pp. 333. \$32.50 (paper).

Jan Lin
Occidental College

Keil's monograph is an impressive contribution to the emerging galaxy of literature on Los Angeles (the "supernova of world cities," declares Mike Davis on a rear jacket blurb), a majority of which are edited collections (e.g., Michael Dear et al., *Rethinking Los Angeles* [Thousand Oaks, Calif.: Sage Publications, 1996], Allen Scott and Edward Soja, *The City: Los Angeles and Urban Theory at the End of the Twentieth Century* [Berkeley: University of California Press, 1996], and David Reid, *Sex, Death and God in L.A.* [Berkeley: University of California Press, 1992]) rather than single-authored monographs (though the offerings of Mike Davis, *Ecology of Fear* [New York: Henry Holt, 1998] and *City of Quartz* [New York: Verso, 1990], are notable exceptions). The utility of Keil's book is the offering of a unified perspective that addresses the city as a whole with less of the redundancy or disciplinary incongruity that may beset multiauthored collections.

Keil's books can be associated with the recent genre of urban studies dubbed the "L.A. school." The moniker putatively distinguishes this repertoire of writing from the proverbial "Chicago school" of urban sociology while asserting the seminal place of Los Angeles in urban studies of the emerging millennium. The L.A. school's theoretical inheritance emanates from the confluence of three principal streams of thought: (a) the French regulation school of social theory, (b) the "world cities" or "globalization" literature, and (c) postmodernism and critical cultural studies. Roger Keil's academic formula can be classified as a variant of the L.A. school enriched by doses of David Harvey and Henri LeFebvre.

As Keil discusses, recent economic restructuring in Los Angeles aptly illustrates the emergence of a "post-Fordist" landscape described by the regulation school. The decline in Fordist mass-production-based manufacturing activities, such as automobiles, steel, and aerospace, has been countervailed by growth in an eclectic array of crafts-based activities such as apparel, food processing, and film production, which rely more on subcontracting, freelancing, and short-batch production methods. Globalization has also wrought its effects, as evidenced in the growth of immigrant participation in garment sweatshops, retail trade in the ubiq-

uitous strip malls and ethnic enclaves of Los Angeles, and foreign investment zones such as Koreatown, Little Tokyo, and Chinatown.

Keil's principal contribution is inserting the centrality of politics into discussions of Los Angeles as the exemplar of the post-Fordist globalizing city. Former mayor Tom Bradley was instrumental in promoting globalization from 1973 to 1986, a period marked by climatic events such as the Olympics of 1984 and the 1986 acquisition of the ARCO Towers building by the Shuwa Corporation of Japan for \$600 million. Within particular subareas of the metropolis, property entrepreneurs manipulated the local political and planning process, sometimes in collaboration with the Community Redevelopment Agency (which commands its own chapter). Globalization is not "moving through cities like an invader from outer space" (p. 146), but evolves as a variegated constellation of growth alliances between local entrepreneurs, state actors, and foreign investors.

The year 1986 was the symbolic apex of an "impossible dream" of sustained growth that concurrently produced the slow-growth initiative Proposition U, which presaged the eventual unraveling of Bradley's multicultural electoral coalition. The L.A. rebellion of 1992 is offered as further evidence of the breakdown of political consensus, a marker of the challenging contradictions of globalization and post-Fordism. These politics are not always chaotic, as Keil highlights the emergence of progressive urban social movements that address labor, community, and environmental justice issues. A seminal example is the multiracial Labor/Community Coalition to Keep GM Van Nuys Open, which bridged old and new left politics.

Keil's book is effective in its discussion of the salience of local politics in mediating dynamics of both synthesis and contradiction under globalization. Less effective is the attempt to define the precise nature of post-Fordist politics and how these politics relate to fragmentation in the production-consumption nexus. The Van Nuys coalition and the renewal of garment industry organizing attest to the continued salience of Fordist-style labor movements. The emergence of environmental justice movements and the Bus Riders Union, by contrast, are indicative of a new politics around consumption and multiracial identity issues. Fragmentation is instead read as suburban balkanization (pp. 173-75) and an outcome of the earlier Fordist era.

Presentation of original statistical data is not a feature of this book. An excellent array of maps, however, help situate the Los Angeles neophyte, as do the copious quantity of photos, almost all taken by Keil himself. An analysis of various films on Los Angeles and the paintings of David Hockney establish the postmodern perspective on the city. Keil does not do a seamless job of stitching together discourses of globalization, regulation theory, and postmodernism in this ambitious monograph, but he has produced an enviable effort on a complex subject.

The Rights Revolution: Lawyers, Activists, and Supreme Courts in Comparative Perspective. By Charles R. Epp. Chicago: University of Chicago Press, 1998. Pp. xv+326. \$45.00 (cloth); \$17.00 (paper).

Sidney Tarrow
Cornell University

Charles Epp's *The Rights Revolution* examines the expansion of individual rights in four Anglo-derived democracies (Canada, Great Britain, India, and the United States) in the late 20th century. Here is its central finding: "Rights are not gifts," they "originate in pressure from below in civil society, not leadership from above. . . . Only certain kinds of pressure from below, particularly organized support for rights litigation, are likely to support sustained judicial attention to civil liberties and civil rights; and support from judicial elites is hardly irrelevant" (p. 197).

These findings are not obvious or casually arrived at; before coming down so firmly on the side of "pressure from below," Epp carefully examines three alternative explanations for the expansion of judicially protected rights in these countries: constitutional guarantees, judicial power, and political culture. He shows that the growth of judicial protection for individual rights has been greatest in the United States and Canada, weakest in India, and present—but not vibrant—in Great Britain. He shows how variations in the presence or absence of constitutional guarantees, judicial activism, public rights consciousness, and robust support structures co-occurred with a greater or lesser expansion of rights in each country.

While the alternative sources emerge as supportive of rights expansion, the most robust factor is "vibrant support structures." With impressive evidence and careful reasoning, Epp shows that, in varying combinations, the existence of public interest groups devoted to rights expansion, private foundation and government support for rights, and—most surprising—an expanded legal profession are the main support structures that connect to the judicial expansion of individual rights.

Epp's study has three major virtues: first, the intelligent use of comparative case analysis to isolate the strength or weakness of equally plausible independent variables in explaining the expansion of judicially protected rights; second, sustained attention to three phases in the expansion of these rights—judicial attention, judicial support, and implementation (a relatively new finding is the importance of support structures even in this postcourtroom phase); third, attention to two important sectors of rights protection that are usually studied in isolation—criminal and prisoner's rights and women's rights.

The Rights Revolution will find its major readership among students of constitutional law and civil rights and in courses in American government. But in approaching problems central to contentious politics from something other than a social movement perspective, it is of fundamental importance to sociologists—and this for three reasons.

First, although Epp's central focus on "support structures" lacks a single citation to the "resource mobilization" tradition in sociology, it lends great support to that tradition: in the key role he finds for social movement organizations, in that of the private foundations that finance these organizations, and in the role of key government actors and institutions in sustaining some kinds of rights campaigns but not others.

Second, that Epp can identify so many "resource mobilization-like" factors in rights expansion in four countries without apparently having heard of the likes of William Gamson, Doug McAdam, John McCarthy, Aldon Morris, the present writer, Charles Tilly, or Mayer Zald may trigger some hubris, but it shows that an intelligent scholar who sets to work on a rich store of data with a plausible theory needs adherence to no school or creed to produce original and robust findings.

Finally, Epp's book is a challenge to sociologists who have defined and operationalized the field of contentious politics narrowly as formal movement organizations operating outside of or against institutions. In finding contentious politics operating at the highest reaches of institutional politics, Epp joins a new phase of study of contentious politics that relates it systematically to political and social institutions. (See, e.g., Anne Costain and Andrew McFarland, *Social Movements and American Political Institutions* [Boulder, Colo.: Rowman and Littlefield, 1998].)

Oddly for a political scientist, missing in Epp's book is a whiff of attention to the political process variables that might have helped him to explain why rights campaigns emerged in some systems more forcefully than in others. And without examining changes in political opportunities and constraints, it is hard to explain why there was so long a hiatus between the creation of a new generation of rights "support structures" in the United States early in the century and the burst of civil rights and civil liberties decisions after 1950.

Nevertheless, Charles Epp's *The Rights Revolution* is a major study in comparative constitutional politics and a significant contribution to and expansion of sociologists' understanding of contentious politics.

Rescuing Business: The Making of Corporate Bankruptcy Law in England and the United States. By Bruce G. Carruthers and Terrence C. Halliday. Oxford: Oxford University Press, 1998. Pp. xv+582. \$60.00.

John Henry Schlegel
State University of New York at Buffalo

In both the United States and England, bankruptcy practitioners and academic experts, observing the rise of corporate/commercial bankruptcies in the 1970s, became convinced that liquidating insolvent concerns was less economically efficient than attempting to reorganize them into smaller, sounder concerns. Because doing so would preserve some, if not all existing jobs, both national governments had reason to agree. And

so, both countries began a process, led at the outset not by professional politicians, but by bankruptcy experts, of reforming their bankruptcy laws. The dust settled—in the United States in 1978 and in England in 1985—with curious results.

Secured lenders found that their security or its equivalent remained, but their ability to demand its sale to satisfy their claims had been compromised. Tax collectors, particularly in the United States, found that not only had their claims been reduced, but also their existing priority among the unsecured creditors. Public utilities in Britain found that they could no longer coerce payments out of their debtors by shutting off utility services. Officers and directors in the United States found that they had the freedom to continue to run their business after insolvency and so had an incentive to seek reorganization, while in Britain these same folks gained no such freedom and moreover faced new liability if they had acted irresponsibly before the insolvency. Consumers in the United States ended up with a modest priority for claims based on deposits made for goods and services that were never delivered, though in England a similar claim and a cognate claim on behalf of unsecured business creditors went nowhere. Attempts to create a national bankruptcy administration in the United States came to naught, while in England an attempt by the government to eliminate such an administration was only partly successful. Bankruptcy practitioners in the United States found that their prestige and income rose sharply with their involvement in reorganizations, while in England a new profession of insolvency practitioner emerged in a complex series of moves that eliminated a large number of the most marginal existing practitioners. And attempts in both countries to rationalize the judicial system's jurisdiction over bankruptcy matters by creating specialized bankruptcy courts failed.

Bruce Carruthers and Terrence Halliday, in an excellent piece of comparative socio-legal-historical research, set out to explain these results. Their careful, predominantly analytical, not narrative, attempt to do so leads to a necessarily long, very interesting book. Two insights are central to their analysis. First, a powerful position in the day-to-day bargaining that is economic life does not necessarily translate into a powerful position when it comes to the bargaining that takes place in the process of legislative reform. Indeed, day-to-day economic strength seems to translate into making oneself a target for others in the legislative bargaining, labeled by the authors as "meta-bargaining," apparently without remembering Leon Lipson's caution, "Anything you can do, I can do meta" (see Arthur Leff, "Unspeakable Ethics, Unnatural Law" [*Duke Law Journal*, 1979: 1299, 1230 n. 2]) Second, changing a profession's jurisdiction, and so the scope of its expertise, is not just a matter of gaining a monopoly market position, as traditional theory seems to hold, but also involves conflicts between professionals in the state, as well as conflicts between state and market actors.

All authors face the problem of convincing the reader of "the truth of the matter stated," as lawyers put it in a different context. The various

norms about the presentation of quantitative sociological data—representativeness of sample, use of appropriate statistical techniques, limitations on causal inference—act as warrants for truth. In historical or more qualitative contexts, warrants of this kind are lacking. What substitutes must a scholar offer instead? At least in work with historical sources, part of the necessary warrant comes from the continual triangulation of the overall narrative with the subparts of that narrative that are used for analytic purposes. Meaning is thus given, and the truth established by the congruence between the narrative apparatus and the analytic one.

In *Rescuing Business*, an extended, detailed narrative is omitted so the analytic pieces dominate, raising in the reader the regular worry that it is the analytics that are driving the authors' conclusions, not the data that would be disclosed by the absent narrative. This worry is unfortunate because my informed guess is that the authors are right on the money in their analysis. One wishes that it were easier to reach that conclusion based on reading the book alone.

Worlds of Production: The Action Frameworks of the Economy. By Michael Storper and Robert Salais. Cambridge, Mass.: Harvard University Press, 1997. Pp. xiv+370. \$45.00.

Gerald Berk
University of Oregon

Worlds of Production tackles a now-familiar empirical problem: How shall we understand the fact that in the post-Fordist world of rapid innovation and trade specialization, national and subnational economies are inserted into the global economy in increasingly diverse ways? Diversity and innovation, the authors demonstrate, are particularly challenging phenomena for existing social theory. Neither neoclassical economics nor neoinstitutionalism (sociological or utilitarian) can secure much purchase on the problem of diversity because they both conceptualize economic selection in ways that tend toward the convergence of economic forms upon a single best practice. Nor do they have much that is useful to say about innovation—the human capacity for collective reflection and choice among possible alternatives—because they conceptualize human action in a deterministic, rather than a creative, way.

Economic theory, therefore, must be reconstructed from the ground up. Drawing upon the philosophy of multiple worlds and conventions, *Worlds of Production* brilliantly deconstructs existing theories and provides an alternative, social constructionist approach to economic action. In this perspective, actors begin neither with the assurance that they alone know their own best goals, nor with internalized social norms. Instead, coordinated action is inescapably social but radically uncertain. Though interdependent, people cannot be confident about the expectation and performance of others. Action then requires shared “conven-

tions," rough discursive rules, or hypotheses about the social and physical world. From the choices actors make about elementary conventions concerning product qualities, technology, and market uncertainty, Storper and Salais classify four ideal-types, or possible worlds of production—market, intellectual, interpersonal, and industrial. Each has its own internal conventions of coherence, innovation, and profitability.

This book then turns to three empirical cases of national trade specialties: fashion and high technology in France, design-intensive manufacturing and precision metal work in northeast central Italy, and high technology regions in the United States. These studies are turned effectively to deconstructive and constructive ends. We learn, for example, how neo-classical institutionalism attains a superficial purchase on the American case because it hypostatizes conventions chosen for particular historical reasons. In three rich narratives, we also learn how real worlds of production are the historical result of deliberations over social identity and forms of participation unique to each society. As such, they are not so much patterned directly upon possible worlds as they are complex combinations. This much, then, explains how nations and subnational regions organize economic innovation in such dramatically different ways.

The role of the state in all this is a complex matter, receiving independent consideration. In an ingenious theoretical move, Storper and Salais reject a priori principles to reimagine state intervention from the perspective of economic actors engaged in coordination in light of their expectations (conventions) about the state. The result is to moot the false distinction between state and society, extend their critique of reified theories of the state, and provide a typology of conventions—"external," "absent," and "situated"—to classify the diversity of relationships between state and economic institutions.

If all this were not enough, *Worlds of Production* has prescriptive ambitions as well. And it is this part of the argument, I think, that raises some of the most challenging theoretical questions. Despite diversity, Storper and Salais point out, not all real worlds perform as well as others, in the sense that they do not create products and processes that give products a viable presence in global markets. If the authors reject all uniform principles of performance (neoclassical or neoinstitutionalist), how do they judge? Success depends on "coherence"—not all conventions conform to one another as well as others. Conventions, they say, cannot be chosen a la carte.

But coherence can be quite slippery in practice. The ideal-types cannot provide the standard because all of the real-world cases are complex hybrids (American high technology, e.g., successfully joins the intellectual to the market worlds), and the book does not provide criteria for successful combination. Moreover, since conventions are formed and revised in a changing global context (itself a social construction), the terms of coherence are likely to be unstable. Finally, as Charles Sabel and Jonathan Zeitlin (*World of Possibilities: Flexibility and Mass Production in Western Industrialization* [Cambridge University Press, 1996]) hypothesize, it

might be that it is precisely the capacity of economic worlds to combine seemingly contradictory forms of action which undergirds their success in responding creatively to change. Nonetheless, Storper and Salais do raise what is perhaps the central challenge for constructivist social theory: Can it provide anything more than nuanced genealogies of social diversity?

All in all, for everything it does accomplish, *Worlds of Production* deserves a great deal of attention not only from students of economic sociology, but also from anyone concerned with the state of social theory. For Storper and Salais not only add a great deal to our understanding of economic adjustment, they provide a model for transcending many of the false dualities that continue to plague all the social sciences.

Tracing the Veins: Of Copper, Culture, and Community from Butte to Chuquicamata. By Janet L. Finn. Berkeley and Los Angeles: University of California Press, 1998. Pp. xviii+309. \$16.95 (paper).

Barbara Sutton
University of Oregon

The age of globalization is hardly conducive to thinking about the meaning of place and community, yet in her study of two mining towns, Janet L. Finn demonstrates that the global and the local are inextricably linked. Not only do global processes contribute to shaping the fates of seemingly isolated communities, but also "local acts and events [can] resist, challenge, and transform the global" (p. 242). This interaction between global and local affairs becomes apparent in the lives and struggles of the people in *Tracing the Veins*. The book presents a historical and ethnographic exploration of culture and community in two mining sites: Butte, Montana, and Chuquicamata, Chile. These two communities existed as "intimate strangers" (p. 6), producing copper for, and at times challenging, the U.S.-based Anaconda Company.

This fascinating journey through copper towns shows with painstaking detail the efforts of mining people to survive and to resist overpowering forces of oppression. *Tracing the Veins* exposes Anaconda's marks on families, local environments, union politics, miners' health, the rhythms of women's lives, and the contours of both communities. Finn skillfully unearths stories of trust and betrayal embedded in both places: stories of broken hopes, broken marriages, and broken promises by unions, governments, and mining companies. She effectively conveys women's struggles to maintain a sense of normalcy in the midst of social crises and their trials as stewards of family and community. Finally, the book contradicts images of structural determinism by depicting people as artisans who craft a life and community with the resources at hand—and in the process engage in social transformation.

Tracing the Veins moves beyond historical, cultural, and political differ-

ences to show the significant patterns that link the two places. As Finn suggests, one cannot "understand Butte's tumultuous past without looking critically at the Anaconda Company's operation abroad, and specifically at the relationship to Chuquicamata" (p. 3). Finn's relational approach allows her to provide a complex picture of the communities and to comprehend the rationale behind Anaconda's business decisions. Thus, we learn why Anaconda boosted production in one place while it sluggishly settled strikes in the other; why the company promoted a partnership facade in Butte while it demarcated more visible social inequalities in Chuquicamata; and how Anaconda's strategic management of production in both sites helped to control copper prices across continents.

A major strength of this study stems from the author's positioning of gender and class relations at the center of the analysis. This useful framework clarifies the inequalities and power contests that pervade mining life and shows how men and women mobilized gender symbols and ideologies as resources for class struggle. While Finn offers a particularly valuable examination of gender at the community level, she fails to problematize the "gendering" of the mining industry. Although women working for mining enterprises are infrequent, this is not reason for their virtual invisibility in Finn's project. Only scant references indicate that some of the workers and union members were women. One wonders what was women's status within the industry, how they negotiated their positions, and whether they ever wanted to or were allowed to be miners. Perhaps, given the hardships and detrimental effects of mining, it was a blessing for women to be spared from this activity. Nevertheless, these issues cannot be taken for granted but need to be looked at critically in order to provide a more inclusive account of mining history. Also, as Finn herself recognizes, this study could be enhanced with a stronger race/ethnicity analysis, especially since ethnic tension did seem to play a significant role in both communities.

Despite these shortcomings, *Tracing the Veins* deserves praise for the nuanced and enlightening analyses disseminated throughout its pages. In the process of unraveling the social history of mining communities, Finn reveals the complexities and contradictions of labor-capital relations, gender relations, and community life. As Finn crosses geographic, political, and cultural boundaries, she also disrupts the borders of academic disciplines. Although grounded in anthropology, this book illuminates problems central to sociology including social conflict and social change, power and meaning, class and gender inequality, social structure and agency, and the impact of capitalism on people's experiences.

Finn resists packing the lives of the people she studied into definite conclusions. Her work hints at the need and possibility of cross-border organizing and engagement "in deep play with the forces that maintain and exacerbate inequities" (p. 246). Her provocative analysis of class and gender urges us to think about relations of consumption along with production and reproduction. Finn offers an eloquent testimony of the value of engaged research by uncovering the dramas, passions, and tensions of

community life in a way that could not have been accomplished by a more distant form of scientific inquiry. Beautifully written, engaging, self-reflective; and insightful—*Tracing the Veins* denotes scholarly expertise as well as sensitivity and compassion.

Manufacturing Ideology: Scientific Management in Twentieth-Century Japan. By William M. Tsutsui. Princeton, N.J.: Princeton University Press, 1998. Pp. xi+279. \$29.95.

Gary Herrigel
University of Chicago

This is a history of the reception, discursive construction, use, and implementation of scientific management in Japanese industry during the 20th century. It rejects both the idea that Taylorism and scientific management are inconsistent (somehow) with the Japanese cultural makeup and the opposite idea that the Japanese industrial system is a more perfect incarnation of the "Fordist" forms of organization that allegedly typify U.S. and European factories. Instead, it claims that ideas of scientific management and Taylorism were in one way or another continually central to debates about factory organization (and often about social and political organization more broadly) throughout the 20th century in Japan. But as in the United States and Europe, producers found that it was not possible to implement the ideas of the master, unaltered in practice. The ideas had to be modified, revised, and compromised in order to be implemented, and Tsutsui insists the problems with implementation and modifications that emerged had as much to do with problems with Taylor's ideas themselves as with cultural particularities of the implementing countries. For example, there was resistance to the more abstract dimensions of the Taylor system and to the relatively heartless, hard sides of the doctrine. Taylorism was, at least rhetorically, humanized in Japan as much as it was in the United States and Germany.

His point is that Taylor's ideas, though powerful and attractive to industrialists, were not perfect and were difficult to implement in all industrial economies—not just in Japan. Tsutsui shows again and again that these difficulties had very little to do with conflicts with indigenous Japanese traditions or cultural understandings. Indeed, he claims that routinely "Scientific Management was smoothly aligned with Japanese conceptions of 'tradition'" (p. 239). More specifically, "Taylorite doctrines—as well as post war imports like Human Relations and quality control—were rooted in "tradition" and almost effortlessly recast as "Japanese-style management" when they were imported and applied in Japan" (p. 240). Tsutsui is so interested in debunking the culturalist reading of Japan that he claims all of the central innovations and distinct features of postwar Japanese production and management organization (just-in-time production, lean production, quality circles, total quality manage-

ment, and the general commitment to productivity within firms and in the society, etc.) are all excellent examples of a "revised Taylorist" paradigm that emerged in Japan at midcentury. The Japanese production system is a fully Taylorized one, according to Tsutsui.

Tsutsui is careful to distinguish Taylorism from Fordism, and he strenuously denies that management and production—and the ideologies about them—are (or were) in any sense "Fordist." For Tsutsui, Taylorism is a rational approach to the organization of production that seeks to separate conception from execution throughout the production process and improve individual worker performance through closer supervision, incentive wages, and more efficient job design (often devised through the use of time and motion studies and statistical review). Fordism, by contrast, is oriented toward the achievement of economies of scale in production by the elimination of labor by machinery. This latter approach was never practiceable in Japan because the smaller, more fragmented Japanese markets meant producers were constantly having to vary what they produced and had to settle for smaller lot sizes. Under such conditions, Tsutsui writes: "The application of Taylorite methods—which promised increased efficiency even to small-scale, labor intensive workshops, and often required little new investment in production technology—seemed to address Japan's industrial ills in a manner that more elaborate Fordist reforms could not" (p. 67).

In this definitional move, Tsutsui distinguishes himself from those who claim that the Japanese system is actually an ultra-Fordist style of regime. By his definition, Japanese strategies simply cannot be categorized in that way. But for all of his efforts to distinguish himself from this latter camp, in the end, his position on how to understand the organization of production in Japan and the quality of work life as a production worker in Japan is far closer to this camp than to those "culturalists" who somehow view work in Japanese factories as involving a more humane and participatory form of capitalism. Taylorism is not Fordism, but it still involves the subjugation of labor by management and the wresting of all residues of worker autonomy and independent deliberation and input from the shop floor. Japanese workplace relations may appear to be more collaborative and participatory, but Tsutsui insists that we should rest assured that this appearance is the same window dressing that culturalists mistook as somehow inconsistent Taylorist ideologies. Group work and multiskilling are not distinctly Japanese cultural artifacts in production that accord labor more input and greater sovereignty in production. They are highly Taylorized organizational forms that Japanese managers devised over the course of the postwar period to take control of production away from workers and improve production efficiency.

This is a clear argument, but it is difficult to know if it is true. Tsutsui mainly defends his Taylorist reading of actual production relations with reference to the evidence provided by the ultra-Fordist writers whose interpretations he rejects. Evidence provided by "culturalist" writers, however, is rejected because they mistake window dressing for substance.

We are never told why one view ruins facts while the other view does not. In the end, one wishes that the same nuance and awareness of the possibility for hybrid solutions that Tsutsui applies so interestingly to Japanese ideological efforts to embrace Taylorism in their own particular context was applied to worker and management efforts to apply it on the shop floor.

Rural China Takes Off: Institutional Foundations of Economic Reform.
By Jean C. Oi. Berkeley and Los Angeles: University California Press,
1999. Pp. xvii+253. \$35.00 (cloth); \$17.95 (paper).

Xueguang Zhou
Duke University

The rapid development of rural industries in post-Mao China presents a challenge to comparative social scientists: Why did the Chinese rural industries take off in a state socialist society with its political authorities largely intact and property rights and incentive structures poorly defined? In this book, political scientist Jean Oi takes up this challenge by developing an in-depth analysis of how evolving institutional context provides both incentives to and constraints on the behavior of local government officials and to induce them to actively promote economic growth within their jurisdiction. Drawing on empirical evidence from both official reports and over 300 interviews in 10 provinces and municipal cities over a span of 10 years. Oi traces the evolution of institutional changes from the early reforms of the 1980s to the more recent changes in the 1990s and shows that local governments played a key role in engineering the rapid economic growth in rural China.

The major innovation in this study is Oi's theoretical framework, which takes rational choice arguments seriously but seeks answers in institutional changes. These changes in the rules of resource allocation and incentive structures induce rational actors to take certain courses of action. Oi asks why local government officials in a communist society are willing partners and even central players in market development that supposedly undermines their own power base. Building on her previous "local state corporatism" model, Oi draws from two streams of ideas to enrich her theoretical explanations. Drawing on the principal-agent literature, Oi focuses on incentives, interests, and information asymmetry (moral hazard) in analyzing the relationships among the central government, local governments, and rural enterprises. By incorporating theoretical ideas from behavioral theories of organizations, Oi pays particular attention to the institutional constraints on the behavior of rational actors such as skills, resources, and available options.

Oi begins her institutional analyses by examining changes in the incentives based on the evolving institutional rules of revenue sharing in the 1980s. She traces two major structural changes: (1) decollectivization un-

dermined the traditional financial resources to local governments and forced them to rely more on collectively owned firms for revenue and (2) the revenue-sharing "contracts" in the fiscal reform gave local governments the "fiscal incentives" to promote economic growth and enlarge their share of the pie, through taxes, within budget and extrabudgetary revenues. Oi details how the pursuit of fiscal incentives led the local state corporatism to evolve in two aspects: to adapt collective enterprises to market production, and to shift more resources to private enterprises.

In this light, Oi argues that "mutual dependence rather than predation is a more apt description of the relationship" between local governments and local enterprises (p. 98). The image is that a local government (e.g., the county government) acted as if it was the headquarters of a large corporation, with township acting as "regional headquarters," and village as the subsidiaries. Using this theoretical framework, Oi examines sources of regional variations in the reform era. She argues that, because of the Mao era's legacy, collective enterprises were in a stronger position (resource configuration, risk sharing) to grow first. This logic also helps explain that variations of collective industries in different regions gave local officials different incentives to adopt different development strategies (chap. 3). The changing financial incentives facilitate an alliance between local governments and local enterprises to promote economic development (including the expansion of the private sector) to serve their own self-interests (chap. 4). This framework also allows Oi to argue and demonstrate various ways that local governments used to protect local enterprises from the central authority (chap. 5). Throughout these analyses, Oi is persistent in taking agents' interests seriously but emphasizing the "institutional foundations" of their actions.

This book captures the most recent institutional changes in rural China. In particular, Oi argues that the interaction between changes in institutional rules and responses by local government officials over time led to an important structural change: the local governments came to act like principals rather than agents, gaining autonomy within their jurisdiction. At the same time, as Oi observes, the economic reform also poses new challenges to the local governments and erodes their basis of power.

I admire Oi's lucid and forceful arguments and think the primary focus on local governments serves her purpose well in demonstrating their central role in rural China's economic growth. But I wish that more attention had been given to the interactions between local governments and rural enterprises, especially from the point of view of the actors in these enterprises. In my view, the intricate interactions among these actors are the key to understanding future directions of institutional changes in rural China.

Since the debate between "the moral economy" and "the rational parent" in the 1980s, social scientists have been struggling with either an emphasis on contextual factors or a focus on actors. Oi shows us that these two are not antithetical. This book demonstrates an excellent blend of substantive institutional analysis and rational choice arguments. Her

study opens a fruitful line of research in studying institutional changes in China as well as other societal contexts.

China's Catholics: Tragedy and Hope in an Emerging Civil Society. By Richard Madsen. Berkeley and Los Angeles: University of California Press, 1998. Pp. xiii+183. \$27.50.

Daniel H. Bays
University of Kansas, Lawrence

In this concise but penetrating and well-written study, Richard Madsen, professor of sociology at the University of California, San Diego, provides a thoughtful essay with a dual focus. He analyzes the recent history and present nature of the Catholic Church in China, and he also discusses the prospects for emergence of a civil society in China, including the obstacles that would need to be overcome and how civil society in China might compare with that of other areas of the world.

Madsen brings uniquely appropriate qualifications to this endeavor. As a former member of the Maryknoll order, he has an acute sensitivity to the dynamics of the Catholic Church. As an experienced China scholar, his decades of observation, study, and writing on Chinese society have given him a deep understanding of the fabric of that society. And his previously authored and coauthored major works on both Chinese society and American society (*Morality and Power in a Chinese Village* [University of California Press, 1984]; *Habits of the Heart* [University of California Press, 1985]; *China and the American Dream, a Moral Inquiry* [University of California Press, 1995]) have explored the role of morality and moral/spiritual values as integrating factors in the fabric of society, giving him the standing as a mature scholar to tackle this large topic in essay form.

Madsen has long had a deep interest in the themes of this book, but only in the early 1990s, as part of a joint research project of the Tianjin Academy of Social Sciences and University of California, San Diego, did the opportunity arise for field research on the Catholic Church in China. Because of the highly sensitive nature of Chinese church-state relations, the Chinese government still restricted Madsen's work, but creativity and the effective use of a very talented Chinese collaborator enabled Madsen to collect an unprecedented (though still modest) quantity of data on Catholic believers today in the major northern city of Tianjin and its rural environs.

One of the results of Madsen's labors is a very insightful profile of the Chinese Catholic Church, which he convincingly portrays as powerfully resilient, especially in its traditional rural core areas, yet also belligerent, contentious, and factionalized. It is in many ways a pre-Vatican II Counter-Reformation church, an historical anachronism. Most of today's Chinese Catholics "have a strong cultural need for hierarchy without the

means to create [it]" (p. 48). They look for vertical authority relationships, not horizontal relationships with non-Catholics. In the world of most Chinese Catholics, loyalty—to the faith, the community, and the martyrs produced by an oppressive government and its decades of persecution—is the cardinal virtue. But this loyalty also contributes to the inability of large parts of the Catholic community to interface constructively with other sectors of society. He provides a particularly insightful discussion of Catholic identity as a form of "ethnicity," that is, Catholicism as ascribed status, with markers such as funeral rituals rather than conscious choice or active religious faith. And he portrays the key element of Chinese Catholicism's stance toward the world around it as one that rejects modernity in favor of trying to retain or reconstruct an all-encompassing "world of God," Catholic social enclaves or bastions of identity that are closed rather than open to the ambiguities of modern economic and social change.

Madsen's analysis of the history and social derivation of the Church today leads logically to rather pessimistic conclusions about the prospects for the Catholic Church playing a significant role in the emergence of civil society in China. Indeed, his short but cogent discussion of the larger issue of civil society unrelated to the Church, and its none-too-sanguine prospects in the context of a powerful state jealous of all potential rivals, is sobering enough. The place where a middle class and the ingredients for civil society might be emerging, the coastal commercial cities, are typically places where the Church is not strong. And if the Church were to play a constructive role, it would need to give much more play to its potential spiritual resources of reconciliation, forgiveness, and healing, as well as social justice, all features of the "kingdom of God" visible in other religious communities but not characteristic of the traditional authoritarian and closed Chinese Catholic "world of God." Madsen's heart is with the visible fragile shoots that might portend such an evolution of the social and theological orientation of the Church. In chapter 4, on urban Catholicism, he bases a discussion of potentially new Catholic social and theological trends on a very thin set of examples. He hopes this might be a trend for the future but also has to acknowledge that there is a very long way to go if the Chinese Church is to transcend the constrictions that its own history has put upon it. Interestingly, Madsen notes that evangelical Protestantism seems to have been more successful than Catholicism in creating faith communities that might have prospects for contributing to civil society.

Forever Foreigners or Honorary Whites? The Asian Ethnic Experience Today. By Mia Tuan. New Brunswick, N.J.: Rutgers University Press, 1998. Pp. xii+203.

Grace Kao
University of Pennsylvania

Forever Foreigners or Honorary Whites examines the ethnic experience of third-generation and beyond Chinese and Japanese Americans in California. Mia Tuan's response to the question posed by the title is that neither label adequately addresses the complexity of the Asian ethnic (those Asians who are third generation and beyond) experience. Tuan's valuable study follows in the footsteps of Mary Waters's *Ethnic Options: Choosing Identities in America* (Berkeley: University of California Press, 1990) and finds that while Asian ethnics practice considerable flexibility in choosing which, if any, cultural traditions of their ethnic ancestry to practice in their daily lives, unlike white ethnics, they have little choice in their primary identification as Asian ethnics. Like Waters, Tuan finds that for Asian ethnics, ethnic identity has more to do with ethnic pride than specific ethnic knowledge or cultural practices. Tuan argues that because Asian Americans are uniquely considered *foreign* by others regardless of their generational status or their cultural practices, Asian ethnics are trapped by the inherent contradictions between their actual identification, as defined by their own lives, and how others identify them. Unlike African-Americans, who are marginalized in terms of socioeconomic status but are not considered foreign (or not American), the Asian Americans she interviewed are assimilated in terms of socioeconomic status and rates of intermarriage but are ultimately seen as foreign, and not American.

Tuan's data come from in-depth interviews with 95 middle-class Asian ethnics living in the San Francisco Bay Area or Los Angeles and found through snowball sampling. This is a respectable sample size, but approximately 25 of her respondents are related. She limits her sample to Chinese and Japanese Americans because these are the only two populations with significant numbers who are third generation and beyond. At the beginning of the book (chap. 2), she acknowledges that this population is atypical of Asian Americans in the United States today, as most Asian Americans are foreign-born (66%, according to her table 2.6, p. 38). However, Tuan does not adequately emphasize that of those Asian Americans who are native-born, a relatively small proportion of them are third or fourth generation. Barry Passel and Jeffrey S. Edmonston (1994) estimate that 83% of the Asian American population in 1990 came from immigrants who entered since 1950 or their descendants ("Immigration and Race: Recent Trends in Immigration to the United States" in *Immigration and Ethnicity: The Integration of America's Newest Arrivals*, edited by Barry Edmonston and Jeffrey S. Passel [Washington, D.C.: Urban Insti-

tute Press, 1994]). Her data are more representative of the experience of Japanese Americans, most of whom are native-born.

This issue of the relative shares of native-born versus foreign-born is an important oversight because much of the experiences of her respondents are directly affected by the large numbers of Asians who are recent migrants to the United States, but it is not a point that Tuan develops. To her credit, she notices that there is tension and resentment toward new Asian immigrants among her respondents. Many of them blame the undesirable cultural practices and habits of their immigrant counterparts on the negative stereotypes of Asian Americans overall. In some ways, they see recent Asian immigrants as the obstacle that blocks them from their entry into the American mainstream. As a reader, I am left wondering whether the experiences of these Asian ethnics would differ if the relative shares of Asian immigrants were smaller. In other words, is the only difference between the experiences of Asian ethnics and, for example, their Italian American counterparts the relative shares of native-born versus foreign-born coethnics? Or, does being Asian offer a unique obstacle to being accepted as American?

Although it is not the focus of Tuan's book, I am left wondering how generational status affects ethnic identity. Certainly there is little difference between the third- and fourth-generation respondents she interviews; yet there are probably considerable differences between the identification of first- and second-generation Asian Americans. Moreover, Tuan skirts the issue of class by examining "middle-class" Asian ethnics, but one is still left wondering how class background affects the likelihood of seeing oneself as an American.

Tuan skillfully paints a rich picture of the contradictions of the Asian ethnic experience and clearly shows that it is not simply like that of blacks or whites. Most, if not all, of her respondents grew up speaking only English at home. They participated in an American life that was stereotypical, complete with baseball and apple pie. Yet, they are constantly asked where they come from, why they speak English so well, why they do not speak Chinese or Japanese, and are expected to have special knowledge about Asia. They feel American, and are as likely to marry same ethnics as other Asians or whites, yet they are still not completely integrated into American society. This is an important book that adds much to our knowledge of the Asian American experience specifically, and the ethnic experience more generally. Tuan moves us one step closer to understanding the complexities of the American ethnic experience that is beyond black and white.

Intimate Ironies: Modernity and the Making of Middle-Class Lives in Brazil. By Brian P. Owensby. Stanford, Calif.: Stanford University Press, 1999. Pp. ix+322. \$45.00.

Miguel Angel Centeno
Princeton University

The middle class has alternatively been the great hope and the preferred whipping boy of social scientists and critics. Sometimes seen as embodying all the best hopes of democracy and economic growth, it is as often castigated for its complacency, political dormancy, economic parasitism, and just plain bad taste. Middle classes have been praised for their contributions to capitalism and democracy and damned for their flirtations with fascism and their insufferable repression.

The middle classes in Latin America have been the subject of much contradictory comment. The continent is often criticized for not having enough of a middle class that would lead it toward a peaceful reform of party democracy and tamed markets. The same middle classes have also been the subject of scathing literary and artistic parodies and have been partially blamed for bureaucratic authoritarianism and the failures of import-substituting industrialization. And yet, we know so little about them. There are many books about factories and plantations, but few about offices; we know about the *favelas* and *pueblos jóvenes* and about the haciendas, but not about the new suburbs. Who are these people who can generate so much paradoxical academic heat?

Historian Brian Owensby's new book on the development of the middle class in Brazil during the first part of the 20th century represents a significant answer. This is a book that should definitely cross disciplinary frontiers, as sociologists without the remotest interest in Latin America will find much here for both scholarly and teaching purposes.

Owensby begins by noting the ambivalent role that middle classes have played in Latin American history and historiography (and for once, I missed the literature review chapter that undoubtedly began this book in its dissertation stage). Owensby nicely contrasts the historical and geographical specificity of the concept of a middle sector with all the universal claims that are made on its behalf. He then proposes a sensible solution to the eternal definitional problems, which, while perhaps leaving more quantitative analysts unsatisfied, reflects a very good reading of Latin American reality. For Owensby, the middle class is made up of that vast population of *gente decente* who wear a tie to work. Distinguished from the rich by their lack of capital, they differentiate themselves from the poor by their abhorrence of manual labor. Add a heavy dose of racial consciousness and exclusion, a fetish for middlebrow culture, and their reliance on a service sector and you have a social phenomenon certainly recognizable to any student of Latin America.

But what is perhaps most interesting in this book is how much the phenomena described by Owensby will be familiar to social scientists

working in other regions. There is, for example, the constant economic and social anxiety that is partly obscured by a sense of moral superiority over the indolent rich and the dangerous poor. As with its counterparts in other parts of the world, the Brazilian middle class occupied the common, if uncomfortable, political middle ground between resenting elite domination and fearing labor agitation. This ambiguity produced frustration and a retreat from politics toward the positivist temptation of administration and expertise. Obsessed with itself and with the future of a modernity, it sought to represent the middle class ends, playing a passive or marginal role in Brazilian society by the 1960s.

Owensby uses a fascinating array of sources to present his well-told tale. Perhaps the most interesting is the series of novels he cites extensively in order to draw a better portrait of the middle-class myth and its disappointments. He augments these with an apparently exhaustive review of newspapers (including an innovative use of classified ads to analyze the job market for clerks in the 1920s) as well as the rich archival records of letters, diaries, and intraoffice memoranda. The final product is an analytically rich portrait of a class caught in so many contradictions as to make its very survival something of an accomplishment.

Owensby does less well when he moves away from the evocative description. The notion of modernity (so linked to the conception of a middle class) is left largely unexplored except in brief syntheses in the book-end introduction and conclusion. He also paints what may be too dark a self-vision of the Brazilian middle class. There is little sense here of the kind of (admittedly misplaced) triumphalism that characterized some of its cultural products and many of its policies. We never quite understand how the middle class contributed to the myth of the brash developmentalist modernity exemplified by the combination of Brasília and *bosa nova*. Also, while doing a good job on the gendered aspects of middle-class life, Owensby pays too little attention to its racial components. This remains, however, a very good book and one that not only teaches a great deal, but also is a joy to read.

Worlds Apart: Why Poverty Persists in Rural America. By Cynthia M. Duncan. New Haven, Conn.: Yale University Press, 1999. Pp. xi+235. \$27.50.

David Brown
Cornell University

Worlds Apart examines the roots of persistent poverty and underdevelopment in rural America during the 1990s. Analyzing data from over 350 in-depth interviews conducted during 1990–95, Cynthia Duncan provides a vivid and highly nuanced description of life in rural America's poor communities. Her analysis of patterns of poverty within and between the Central Appalachian coalfields, the Yazoo Delta of Mississippi, and a

remote resource-dependent area in northern Maine tells us much about what it is like to be poor in a poor area in contemporary rural America. Her analysis demonstrates that where one lives matters. Strikingly similar social structures in Appalachia and the Delta constrain opportunities for the poor, while a distinctly different organization of social relations in northern Maine facilitates social mobility. This book contributes to our understanding of the causes of persistent poverty, and it demonstrates that rural poverty is a persistent and intractable problem in America today. This work clearly demonstrates that "welfare to work" policies have a better chance of raising poor persons out of poverty in some environments than in others.

Both Central Appalachia and the Yazoo Delta are characterized by rigid two-class systems and corrupt, nondemocratic patronage-ridden politics. In both instances, the haves and have nots are segregated from each other in education, housing, the church, and voluntary organizations. Moreover, control over public- and private-sector jobs is intertwined with property ownership in both areas, and hiring is by "invitation only." While the poor and nonpoor are socially isolated from one another, they are also economically interdependent. Since the haves control access to jobs, education, and even the social welfare system, they are able to maintain the poor in a dependent position. Norms of interclass (and interracial) behavior are clear and visible, and sanctions for violating these norms can be severe. Finally, since the haves benefit from maintaining the status quo, they have no incentive to contribute to strengthening local institutions or seeking economic development. In contrast, the low-income community in northern Maine studied by Duncan is characterized by norms of inclusive participatory governance, a more even distribution of income, and benevolent local business leaders who invest in the community. Hence, this community has a rich civic culture and strong social institutions that promote pathways from poverty for the poor.

Duncan does not make an independent contribution to sociological theory, but her analysis is shaped by at least three strains of contemporary social thought on the causes of poverty. She demonstrates that Wilson's explanation of the causes of urban poverty is also useful for explaining the persistence of extreme deprivation in rural areas. Poor men and women in Appalachia and the Delta are disadvantaged by a lack of jobs and the same kind of social isolation that Wilson describes in parts of urban America, while social buffers and effective community institutions in northern Maine help the poor overcome the constraints inherent in a resource-dependent rural environment. Duncan's work also helps to bring culture back into the analysis of the determinants of persistent poverty.

Using Swidler's concept of culture as a tool kit, not a set of values or preferences, Duncan argues that poor persons behave differently in different communities because they learn different skills and lessons in these different environments. And finally, Duncan's work contributes to our understanding of how local social capital determines life chances. Duncan

argues that in northern Maine strong social capital contributes to alleviating poverty, while in Appalachia and the Delta, the lack of inclusive civic norms undermines the poor's life chances.

I am enthusiastic about this book, and I recommend it highly. I would be remiss, however, if I did not indicate several concerns that one should bear in mind while reading it. Duncan acknowledges that locality-specific poverty is a historically contextualized process, yet her treatment of the history of class and race relations and of underdevelopment in Appalachia, the Delta, and northern Maine is superficial at best. Duncan's study also suffers the same shortcoming that could be lodged against any comparative case study, for example, one wonders how far to generalize the conclusion that social capital is a principal determinant of social mobility among the poor. Broad theoretical statements such as this are best tested with a combination of in-depth qualitative study and multilevel quantitative analysis of large-scale panel data.

The Silent War: Imperialism and the Changing Perception of Race. By Frank Füredi. New Brunswick, N.J.: Rutgers University Press, 1998. Pp. 282. \$50.00 (cloth); \$20.00 (paper).

John Willinsky
University of British Columbia

The challenge of postcolonial studies is to convince readers to move with great deliberation beyond what is largely forgotten. Yet the legacy of imperialism can hardly be overcome if we fail to realize what it made of the world, and for that reason, we should be grateful to Frank Füredi for his ongoing history of its ideas and ideologies. In *The Silent War*, Füredi focuses on the imperialist role race played in the international arena during this century. Imperialism's unmistakable weight is felt in this work, as the imperial powers projected their anxieties over declining empires onto threats of race war and racial integration.

Füredi makes disturbingly clear how these racial concerns played themselves out in Western policies throughout the century, with a special focus on the two world wars and the Cold War. Where his previous book, *New Ideology of Imperialism: Renewing the Moral Imperative* (Pluto, 1994), dealt with contemporary Western responses to Third World nationalism, this new one explores the West's historical linking of such concepts as "race consciousness" and "nationalism" as an impetus for constraining and quieting such expressions of resistance to imperialism. Having scientifically fashioned this monstrous concept of "race" during the last century, in an effort to place imperialism within the natural order, the West had then to struggle to keep it from becoming the source of its own moral and political undoing.

Yet was it really a "silent war" over race, as Füredi's title suggests? I was not convinced this was the right phrase for it at all. It takes two to

war, and Füredi tells the story so thoroughly from the perspective of the ruling white nations that it is hard to envision how this race consciousness was actually working among the colonial peoples. Thus, we lose a sense of their agency and a gauge for assessing the West's response in the successful struggle against imperialism.

Rather than a war, what Füredi is naming is something equally insidious—"the way in which [white] racial fears have shaped the discourse of race" in ways that reflect its own "hesitancy, bitterness, scepticism, bad faith and above all the desire to slow things down" (pp. 234, 236). That is, this book is about how this pervasive and unsettling race anxiety led to suppressing or silencing what had become of this concept of race while continuing to use it as affirming imperialism's natural order. Füredi brings to light the official expressions of concern, as well as the newspaper editorials, on the arming and training of colonial natives in the two world wars; the rise of Japan as an imperial, anticolonial military power, which was engaging the West in what was seen as an initial "race war"; and the rise of race-based nationalism within the colonies.

A second major contribution of this book is its convincing portrayal of the links between domestic and foreign policies over race, again in the strategic play for power in the guise of responding to a shifting but natural moral order. Füredi opens our eyes to how the Civil Rights movement successes and support from the U.S. federal government were part of a Cold War strategy to compete against Soviet propaganda attacking the moral claims of a racist West. It was not so much the horrible lessons learned about racism from the Holocaust that drove civil rights support, in Füredi's analysis, but the West's interests in competing against the communist block and securing Third World allegiance. These attacks on racism, too, were cast as falling within the scope of this moral and righteous crusade, this white man's burden without end. The extent to which race and equality were no more than a shell used in this international struggle can make for discouraging reading, as one realizes how those on top are ever-vigilant, ever-anxious with their power, using whatever concepts are handy, while at least never feeling completely assured in their efforts.

Yet within this Western anatomy of power, we would also do well to study the academic contributions that touch on our own field. Füredi turns, for example, to the celebrated anthropologist Bronislaw Malinowski, in his concerns during the 1930s and 1940s with the "tragic configuration of racial relationships," published on occasion within the pages of this very journal (p. 150). The academic stance proved only slightly more progressive than the racist press in advocating the preservation of racial purity. Social science research on "oppression psychosis" and "hyper self-consciousness" among colonial intellectuals succeeded in blaming the victims and making them seem crazy for taking this race thing so seriously. The literature on "race relations," Füredi also makes apparent, largely dealt with the white response to race consciousness. This is part of what makes this book a guide to our own times, when we are again looking

for colorblind policies and programs, "as the sentiment of racial anxieties has become even more silent" (p. 238).

Despite a few repetitious points in the book and places where I would have liked to have had the wealth of impressions and concerns expressed by officials in memos and policy documents corroborated and challenged by other related sources and nonwhite perspectives, these do little to overshadow Füredi's accomplishment with this historical project on imperialism. It has much to offer postcolonial studies in the social sciences.

Why Americans Hate Welfare: Race, Media, and the Politics of Antipov-erty Policy. By Martin Gilens. Chicago: University of Chicago Press, 1999. Pp. xii+296. \$25.00.

Sanford F. Schram
Bryn Mawr College

This book's title is surely a loaded question. Yet, that is Martin Gilens's point: by the time Americans are asked about welfare, it has been distorted beyond recognition. In the end, so is welfare policy. How does this happen? In this meticulously researched and well-written book, Gilens documents how many Americans come to misunderstand welfare and how those misunderstandings influence welfare policy making. For Gilens, a prime culprit is the news media. Moving from portraits of public opinion to analyses of news reporting to finally a critical examination of the welfare reforms of recent years, Gilens adroitly uses a variety of empirical sources to demonstrate that (1) Americans are supportive of government aid to "deserving" recipients; (2) they are however misinformed about who is receiving assistance because of how the news media misrepresent welfare recipients; (3) visual misrepresentations are especially influential and misleadingly overrepresent African-American, single mothers as recipients; (4) journalists and editors in particular are almost as misinformed as the general public about the real proportionate breakdowns of the welfare population as well as the extent of their own misreporting; and (5) race is systematically related to the making of welfare policy, with welfare being considered a "black" program that therefore is less deserving of support.

What an indictment, especially given the methods and data used. Gilens creatively analyzes public opinion polling data to illuminate mass support for not just the welfare state in the abstract, but also for welfare to the poor in particular. Gilens reviews his already widely noted, previously published research on racial attitudes and opposition to welfare. This book however goes well beyond the survey data. It builds on other work Gilens has done on news media representations of welfare recipients. In perhaps the most interesting section, there is an analysis of how media representations of people living in poverty have changed over time. Gilens finds that overrepresentation of African-Americans in stories

about poverty started to increase in the 1960s when rapid growth in the number of welfare recipients occurred under conditions of political instability, racial turmoil, and civil unrest. Gilens notes that the welfare population both black and white grew during this period in proportionate amounts, maintaining whites as the largest group. Yet, the public increasingly came to see welfare as a program for African-Americans. Gilens finds that the news media have for years now overrepresented African-Americans in negative stories about poverty and welfare and that these are tied to the growing mass antipathy toward welfare.

Gilens not only makes a real contribution to understanding the role of mass communication in the making of social welfare policy, his work here adds significantly to the growing body of research that demonstrates that the news media do not simply reinforce existing political attitudes but change them as well. Gilens's analysis eventually turns to the role of race in the welfare retrenchment of 1996. He helps us understand how the public could come to support draconian welfare reforms even as it opposed cutting back assistance levels to those who are genuinely needy. With careful attention to the particulars of welfare reform, Gilens reviews survey data to suggest that Americans supported the welfare retrenchment of 1996 based on the mistaken assumption that most welfare recipients were underserving, that is, were not trying to achieve personal responsibility in regards to work and family.

Yet, it is here and in its conclusion that this excellent book falls short. Gilens is too content to accept as legitimate that the American public relies on the prevailing distinction between the "deserving" and "undeserving" poor. He faults the news media not for reinforcing this problematic distinction but only for misrepresenting it. The implication is that if the stories had been accurate, that is, if most adult welfare recipients were African-American, single mothers who chose welfare over work, then the public's opposition to welfare would be justified. Gilens's implied solution is to counter these false images with more correct images of recipients as people from a variety of ethnicities who are trying to become self-sufficient. This is not enough. It leaves the false dichotomy of merit intact, allowing it to continue to marginalize poor single mothers who need public assistance in lieu of the inadequate support that paid employment might provide them. Simply put, Gilens fails to consider the possibility that no one "deserves" to be poor. If he had done so, he could go beyond emphasizing that the media exaggerate the extent to which African-Americans rely on welfare to the more critical issue of how the political economy systematically works to relegate disproportionate numbers of African-Americans to lives of poverty and deprivation.

Gilens effectively uses the "master's tools" in the form of quantitative analysis of public opinion data, but without enough of the necessary critical reflection. He recognizes these tools can only take us so far; however, his limited conclusion suggests we need to go further.

Postmodernity and the Fragmentation of Welfare. Edited by John Carter. London: Routledge, 1998. Pp. 294.

C. G. A. Bryant
University of Salford

Social policy teaching in British universities has typically been located in separate departments of social policy or semiautonomous social policy sections of joint departments of sociology and social policy. These have sometimes served to insulate it from theoretical debates that have enlivened/bedeviled other social sciences. Hence the accusation—often unfair—that the academic study of social policy has been atheoretical, even empiricist. It is widely recognized, however, that it cannot so remain because the original rationale for social policy—the application of reason to the public determination and delivery of welfare, via the state and other agencies, to all citizens in a context of economic growth and social progress (a very grand narrative indeed)—has been undermined by economic, social, and cultural developments in the 1980s and 1990s. The age of Enlightenment has finally closed. Simple confidence in reason, science, modernity, and progress, and in the state as the prime instrument for the delivery of welfare, has been lost. Contributors to this volume are not always secure in their characterization of recent developments—post-industrialism, postmodernism, poststructuralism, late modernity, postmodernity, post-Fordism, globalization, and neoliberalism are all invoked and variously defined—but they are clear that the articulation and execution of social policy are enormously more problematic than was the case even a generation ago. It is notable that the theoretical inspirations for their reformulations are less Titmuss and Marshall than Foucault, Giddens, and Beck.

The volume originated in a conference at the University of Teesside in September 1996. The papers collected here were revised in early 1997, that is, before and after Labour's triumph at the general election of May 1, 1997. Tony Blair's (and Tony Giddens's) "third way(s)" may not have been presented as such until the summer of 1998, but contributors to this volume largely saw them coming. It offers as comprehensive and engaging an overview of contemporary issues and dilemmas in social policy *as perceived in Britain* as anyone could reasonably wish for. Even the sparing reference to the European Union (with the exception of Ginsburg's chapter) is telltale. The editor's introduction is followed by 16 chapters in five parts. In part 1, Carter, Gibbins, O'Brien, and Penna, and Fawcett and Featherstone discuss "postmodern frameworks and social policy." Part 2 consists of Mann, and Croft and Beresford, on "critical social policy and postmodernity." Next, there is Carabine, Smaje, and Nettleton and Burrows on "social divisions and social exclusion." In part 4, Clark, Ball, Bunton, and Loader deal with "governance and new technologies of control in the new social policy"; and in the concluding part, Hoggett

and Thompson, Cochrane, and Ginsberg consider "citizenship amid the fragmented nation state."

Every one of these papers is worthy of comment, but I only have space here to mention six. My choices are those of someone with special interests in social theory and political sociology. (1) Gibbins's discussion of postmodernism and poststructuralism combines lucidity and concision better than any other I have seen—including essays by social theorists. (2) Mann engages with Bauman and Giddens to begin to make clearer what a welfare politics beyond left and right might be like. He points out that Giddens's emphases on risk and risk assessment bear upon the concerns of both "orthodox" and critical approaches to social policy. (3) Nettleton and Burrows connect quite separate literatures on reflexive modernization and the political economy of health and housing to indicate how we no longer seem able to recognize and address real need. "Surely it can only be under postmodern conditions that policy interventions based upon providing more affordable and decent housing could have become almost unthinkable? (P. 164.) (4) Clarke is interesting on the displacement of bureau-professionalism by managerialism and its connection to the changing, dispersed form of the state. "The state delegates its authority to act in specified ways to subaltern organisations. Dispersal does not carry the intrinsically disintegrative connotations of fragmentation" (p. 176). Indeed, managerialism is expected both to bring street-level bureaucrats under control and impose order on the chaos of the new. (5) Hoggett and Thompson assess an associationist vision of welfare delivery derived, above all, from Hirst. They argue that the pros outweigh the cons, but the fatal flaw of associationalism may turn out to be its utopianism (in the sense of Marx and Durkheim)—it is hard to see by what steps it could be brought into being. (6) Ginsburg is right that Delors's modernist ambitions for a Social Europe are no longer relevant given that the European Union is not evolving as a suprapstate. More questionably, he does not expect the EU's concern for social cohesion to translate into effective social policies because it lacks the means to implement them and such policies are in any case incompatible with commitments both to liberal economic reform and Economic and Monetary Union (EMU). It is disappointing that other contributors have paid so little attention to the EU.

States, Markets, Families: Gender, Liberalism and Social Policy in Australia, Canada, Great Britain and the United States. By Julia S. O'Connor, Ann Shola Orloff, and Sheila Shaver. Cambridge: Cambridge University Press, 1999. Pp. ix+281. \$64.95 (cloth); \$24.95 (paper).

Lisa D. Brush
University of Pittsburgh

For the past two decades, researchers have sought to bridge the yawning gap between mainstream and feminist welfare states studies. With their new book, Julia O'Connor, Ann Orloff, and Sheila Shaver contribute the capstone to this endeavor. Their conceptual framework, analytic strategy, and comparative data give pride of place to the gendered character of—and connections among—states, families, and markets.

At the core of their contribution is their focus on "social policy regimes—patterns across a number of areas of policy" (p. 12). Proponents of regime analysis (e.g., Gøsta Esping-Andersen, *The Three Worlds of Welfare Capitalism* [Polity Press, 1990]) group under the rubric "liberal" the industrialized democracies committed to limited state intervention in both markets and families (Australia, Canada, Great Britain, and the United States). O'Connor, Orloff, and Shaver assess their grouping by analyzing the similarities and differences in the origins, outcomes, and restructuring of gendered welfare policy in these four countries. They build on the concept of policy regimes and on the analysis of the feedback loops between state institutions and policies, on the one hand, and central features of stratified civil society, on the other.

The empirical heart of the book beats through three chapters, one on labor markets, one on income transfer programs, and one on what the authors call "body rights"—that is, "the role of the state in the sexual and reproductive lives of its citizens" (p. 157). They show how specific constellations of institutions, political practice, and rhetoric shape policy regimes. One important payoff of their detailed analysis is the revelation of the extent of variation in gender outcomes within liberal welfare states. The authors show how cross-national differences in strategies, institutions, and ideologies have significant consequences in struggles for gender equality as well as noteworthy effects on policy outcomes.

Mainstream comparative welfare state studies frequently naturalize assumed differences between women and men and then place a seemingly "universal" subject (actually a masculine worker or citizen) at the center of policy analysis. The authors "gender" this approach by accounting for the specificity of women's experiences (which they acknowledge are diverse). They also reconfigure the concepts of mainstream welfare state studies (citizenship, rights, interests) using a notion of gender as a form of social organization embedded at the core of state institutions and practices. They refuse to confine gender to families and interpersonal relations, and elaborate the many ways class, race, and gender interact to

reproduce welfare states and social policies as “mechanisms of stratification” (p. 232).

O'Connor, Orloff, and Shaver are not preaching to the converted. Their intended audience is the mainstream of political sociology. The authors are deeply engaged with the concepts and methods of mainstream comparison. They provide ample, credible evidence of the centrality of gender to welfare state outcomes and processes. The book therefore should persuade those previously unconvinced of the value added of a gendered analysis of welfare states. It provides a wealth of data and a useful framework for focusing the research that remains to be done. It will be immensely useful for informing nonspecialist scholars about the substance and stakes in the debates over gender, class, and social provision. It also shows graduate students exactly what it takes to do gendered welfare state studies.

Actually, it is a good thing that the authors are not preaching to the choir, because the book is unlikely to move feminists to hymns of praise. Their determination to reshape the mainstream means the authors reproduce some of its common problems. The authors reduce women's sexuality to reproduction. Their research on “body rights” centers almost exclusively on abortion, an important case of intrusive state interest in women's bodies. Because they do not analyze the gendered liberal view of sexual personhood, however, their critique never assesses the contradictions feminist analysts confront (e.g., Linda R. Hirshman and Jane E. Larson, *Hard Bargains: The Politics of Sex* [Oxford University Press, 1998]). While they focus on women as workers, mothers, and caretakers, they only tangentially discuss women and men as policy makers, power brokers, ideologues, or strategic apologists of the welfare state. Their focus on “autonomy” seems oddly disconnected from the realities of many citizens' restructured and downsized lives, especially those of poor women who are the most frequent targets of stigmatizing programs and vitriol in liberal welfare states. They only gesture toward violence against women and the comparative research that implicates it in state power, social policy, and women's second-class citizenship. They interpret the gendered character of the state in the context of a diverse set of political and economic factors that nevertheless excludes masculine power or privilege (pp. 37, 235). The point is, however, that these three scholars know their audience and have written a book that should move quickly into the canon of welfare state studies, with transformational effect.

What Difference Does a Husband Make? Women and Marital Status in Nazi and Postwar Germany. By Elizabeth D. Heineman. Berkeley and Los Angeles: University of California Press, 1999. Pp. xviii+374. \$45.00.

Lynne Haney
New York University

Over the last decade, feminist scholars have been engaged in an interdisciplinary dialogue about historical and cultural variation in the category of gender. Historian Elizabeth Heineman adds depth to this dialogue in her outstanding book, *What Difference Does a Husband Make? Women and Marital Status in Nazi and Postwar Germany*. Through a brilliant study of German women "standing alone," Heineman advances a provocative argument: marital status operates as a social marker analogous to race, class, and gender. Like other categories of difference, marital status is fluid and constructed by states as they develop. And, like other categories, the meaning ascribed to marital status has profound implications for women. Heineman's book explicates these implications for German women as they passed through Nazism, Communism, and liberal democracy.

The story Heineman tells is breathtaking in its depth and scope. The narrative begins with the rise of the Nazi regime, which was quite invested in policing the borders of marital status. The Nazis codified a three-tiered classification: "Asocials" who were not worthy of marriage; worthy, yet-to-be-married women who had access to paid employment; and already-married women who were to remain housewives. These divides began to erode with the onset of the war. With the deportation of asocials and men's departure for war, the lines between the unmarried and married blurred. But they did not disappear: war wives and single women continued to have differential access to state support, employment, and sexuality. Marital lines were blurred further in the years of military defeat and occupation. Amid extreme hardship, the experiences of "women of the rubble" were shaped less by marital status and more by their survival strategies.

A major strength of Heineman's book is thus its compelling account of shifts in the significance of marital status as Germany moved into and out of war. This account becomes even more powerful in Heineman's synchronic analysis of East/West developments in the postwar era. The divergence is striking: both regions came out of the war with the category of marital status unstable, yet the divide was reinstated in the West and accrued less significance in the East. West German policies and public discourse reconstituted single women as a distinct social group; the re-vamping of civil codes, welfare provisions, and labor regulations imbued wifehood with renewed legal protection and security. By contrast, the Communist regime in the East narrowed the social distance separating the unmarried and the married. Here too there were profound legal and economic changes, but these shifts basically pushed all women into wage

labor. Although married women received special benefits at work, the discourse of wage labor gave both unmarried and married women resources for claimsmaking. This argument about regional difference will be of particular interest to comparative state scholars; it affirms the insights of scholars like Gøsta Esping-Andersen and Ann Orloff about the institutional underpinnings of different state regimes. And Heineman's argument about marital status as a central, albeit fluid, form of differentiation adds complexity to feminist work on the construction of gender.

Of course, one could argue that the larger relevance of Heineman's arguments is limited by the specificity of her case: it is not entirely surprising that the meaning of marital status varied in Germany given the social upheaval of war. Heineman does not reject this explanation; she links changes in marital status to the demographic changes of war. At the war's end, there were 7 million more women than men living in Germany. In such dramatic circumstances, the meaning of marital status had to change. Hence, while Heineman's story is a demographic story, it is not *only* a demographic story: Demographic changes still had to be interpreted and addressed.

It is here that Heineman reveals what difference a *state* can make. Heineman shows that states make choices and decide how to deploy marital status. In Nazi Germany, marital status was policed through the work of Nazi officials, social workers, and physicians. In the postwar period, the liberal state's policies widened the unmarried/married divide, while the Communist state's ideology narrowed it. While convincing, this state analysis left a few lingering questions. Those familiar with socialist Eastern Europe may wonder why Heineman relies so much on state ideology and so little on political economy. Given the huge gaps between ideology and reality, a fuller account of actually existing state socialism may have added nuance to her East/West argument. In addition, while Heineman proves that the state makes a difference, it is not entirely clear why. What explains these different state politics of need interpretation? Was it all ideological? Or was it linked to state structure and organization? There is a powerful theoretical argument about the state lurking in Heineman's historical account; had the analysis been more firmly grounded in the comparative literature on gender/state regimes, this argument would have blossomed.

Similar questions surface in Heineman's analysis of the role of marital status. While Heineman convincingly demonstrates that husbands do make a difference in women's lives, it is unclear how this insight should be incorporated into feminist scholarship. Under what conditions does having a husband accrue particular political significance? When does standing alone have its most severe material consequences? Connected to this, How does marital status relate to other categories of difference? Heineman's historical narrative reveals complex intersections among marital status, race, gender, and class: the Nazi regime connected marital status to its larger racist/eugenics project; the postwar liberal regime often used it as a proxy for gender; and the postwar Communist regime

linked it to a logic of class. So what do these uses of marital status reveal about the relationship among categories of difference? Here too, there is a profound theoretical argument about interlocking classification schemes underlying Heineman's historical analysis. A fuller engagement with feminist theory and social science could have brought this argument more clearly to the fore. Thus, Heineman's book not only provides a superb account of the social construction of marital status, but it also testifies to the importance of maintaining interdisciplinary feminist dialogues.

At the Breast: Ideologies of Breastfeeding and Motherhood in the Contemporary United States. By Linda M. Blum. Boston: Beacon Books, 1999. Pp. 227.

Barbara Katz Rothman
City University of New York

One of the things I really love about being a sociologist is being forced to continually call into question every bit of taken-for-granted reality, every treasured belief, every deeply held value. I had to keep reminding myself of the joys to be had in that as I read Blum's wonderful, challenging book. I just cannot seem to maintain that ironic distance when I look at breastfeeding, one of the deepest, truest, most satisfying parts of my life. Nursing three children, two by birth and one adopted, each way past any age normally considered acceptable in America, was an experience that helped define me. But what did it all mean?

Blum opens with a page of four epigraphs, one from Patricia J. Williams from *The Alchemy of Race and Rights*, asking: "Is there not something unseemly . . . about the spectacle of a white woman mothering a black child?" I have spent countless hours staring at my black child suckling my white breasts, and I have tried and tried to understand what I was seeing. Generosity? Transcendence? Or white imperialism?

Nothing about the breast is self-evident. Nothing is self-evident, the sociological insights with which I began and which Blum embraces. But the breast is particularly problematic—deeply sexualized, deeply racialized, deeply contested. Who owns it, the tug of war over the breast, occurs all over this culture: between husbands and wives and two-year-olds not-yet-weaned; between surgeons and art directors and lingerie designers and people wearing pink ribbons and mammographers and pimps and mamas.

The resurgence of breastfeeding comes in the midst of all of this. It is, as Blum shows, a class-specific, culturally based resurgence. Those of us in the coalition of social movements that urged a return to breastfeeding—feminists, women's health activists, traditionalists, child health workers—were mostly white, mostly middle class, mostly comfortably placed and mostly, Blum reminds me forcefully, oblivious to all that. When we said that breast was best, that breastmilk was for babies, cow

milk for calves, it never, ever occurred to us how it was going to be used against women.

And just how has it come to be used against us? Here's a hint: a baby suckling at a breast is not spending any money. Not only that, a capitalist sin of the first order in itself, but waves are being lost as the mother, a possibly valuable middle-class educated worker, sits in a rocking chair rather than at her desk. Enter the solution: buy a breast pump, bottles, nipples, all of the paraphernalia of bottle feeding, pump the milk and hire a less expensive caregiver to feed the baby its mother's milk. Voila! Another problem solved.

It has been relatively easy to sell white women on this solution: there is the appeal of the "natural." The sense of "doing the best" for your baby, and the absolutely uncanny accomplishment of being in two places at once: part of your body, your milk, is at home with your baby while the other parts are at work. Blum has interviewed women who express their acceptance of this arrangement, this—as she explains—fetishizing of the breastmilk as the mother is erased. It is not a perfect solution, but it works, more or less. As does breastfeeding itself—the women Blum interviews are more often presenting stories of one or another kind of breastfeeding failure than of breastfeeding itself. Medically managed births are not conducive to breastfeeding, and Blum documents this with her interviews.

Nor are, as Blum carefully shows, the lives of women well-suited to breastfeeding in America. A "pumping station" is a cheaper corporate solution to the "problem" of motherhood than decent maternity leave would be. And for women who are not in elite work, even those small concessions are unavailable. Privacy, time, respect, time, support, time—the things one needs to suckle a baby—are just not there.

The pump-and-leave approach is a harder sell for women of African descent. One of the great contributions Blum makes—and a good reason to assign this book in research methods classes, classes on race and ethnicity, as well as on gender, medicine/health and family—is her thoughtful and appropriate use of race. "The natural" is a less persuasive argument for women with "such a long history of oppression justified by their closeness to nature, their primitive, subhuman being" (p. 14). Given the larger historical context of black "mammies" suckling their white owners, and the immediate personal context of intrusive medical monitoring during pregnancy, it is a small wonder that women of African descent "wanted the state off their bodies . . . wanted their bodies back for themselves!" (p. 176).

At the Breast is feminist research of the highest order, setting a standard for how the work ought to be done. In addition to her systematic reading of the varied literatures of breastfeeding, she did a bit of participant observation with La Leche League (LLL) meetings, and then 24 interview conversations with LLL mothers. To get "the other side," women not so committed to breastfeeding, and women of lower SES, she did fieldwork in an urban family practice clinic, 34 interviews with moth-

ers in that clinic, and then 19 interviews with low-status service employees at a large hospital, where she found most (17 out of 27) of her black interviewees. What is striking and admirable about her analysis is that in a discussion that has been almost entirely subsumed in concerns about what is best for babies, Blum focuses unwaveringly on mothers, on women as minded social beings. She explicitly links infant feeding choices and practices to questions of citizenship, national and global politics, and shows that "as such the meaning of the maternal body should not be seen as the periphery of scholarship" (p. 201). These are questions "fundamental to social justice" (p. 201); in the words Blum uses to begin her last paragraph: "It matters" (p. 201).

The Body in Everyday Life. Edited by Sarah Nettleton and Jonathan Watson. London and New York: Routledge, 1998. Pp. x+308.

Barry Glassner
University of Southern California

The editors of this collection achieve their modest but estimable goal of bringing data to bear on what has been a highly theoretical field of inquiry. Sociological work about the body suffers "from theoreticism, a condition which implies that attention is limited to theory, which in turn is not grounded in the empirical domain" (p. 2), Sarah Nettleton and Jonathan Watson argue in their introduction. By means of the 14 qualitative studies they chose for the volume, they intend not to resolve or transfigure theoretical disputes within the sociology of the body, but rather "to inject the contemporary debates with some more grounded material" (p. 3).

Nettleton and Watson go too far when they claim that the existing literature contains few studies in which people speak about their bodies—on my shelves are many dozens of books and papers in which such talk occurs. Most of that work has come out of the United States, however, and in Great Britain, where all but two of the contributors to this book reside, empirical research on embodiment indeed appears to have been rare. In Britain, perhaps more than in the United States, G. H. Mead's observation still applies: "We are trained by our society to keep our bodies out of our minds" (quoted on p. 145).

So this collection of small-scale studies—most rely upon interviews or observations with anywhere from a dozen to several dozen people—does help to fill a gap in the sociological literature. The authors consider embodied activities ranging from breastfeeding to bodybuilding, and they pay careful attention to the ways in which beliefs about the body and experiences of pain and pleasure are gendered class specific.

A few of the papers are all the more welcome for going beyond the accustomed approaches in grounded theory research for the collection and interpretation of data. For instance, Simon Williams and Gillian Bendelow not only conducted interviews with children, they also asked

them to draw pictures about cancer. The resulting drawings, some of which are reproduced in the book, vividly reveal the demonic imagery of the disease within the culture. Some of the children drew bodies that the authors appropriately describe as having "a chilling, Kafkaesque/nightmarish quality" (p. 110), and others drew disfigured bodies with parts missing. The children's encounters with cancer were not limited, however, to frightening imagery. Many of them demonstrated a great deal of knowledge about the different varieties of cancer or causes of the disease.

Another contributor to the volume, Paul Higate, took an entirely different and equally innovative approach. In order to investigate the ambiguities of masculine embodiment within military contexts, Higate made use of autobiographical notes he took while working as a clerk for the Royal Air Force. Clerks must negotiate, Higate shows, contradictory expectations of themselves as strong soldiers and as presumptively effeminate because they do secretarial work. Through a careful analysis of how body differences such as height and build get put to use in negotiating power relations in this setting, Higate gives substance to Chris Shilling's provocative notion of the body as an "absent presence" in persons' experiences and in social interactions (*The Body and Social Theory* [Sage, 1993]).

In her contribution to the collection, Alexandra Howson departs in still another productive way from what has become a convention within the sociological study of the body. Howson takes on a core Foucauldian assumption. In a paper on cervical screening practices, Howson challenges the heavy emphasis on medical practices and medical surveillance (in Foucauldian terms, the medical gaze and the Panopticon). By focusing on women's experiences of self-examination and their encouragement of friends and kin to go for cervical screening, Howson is able to demonstrate that women do not necessarily experience screenings as disembodied or disempowering. On the contrary, for many of the women Howson interviewed, screenings were seen as either entitlements or as conscious efforts at self-maintenance.

Howson's findings underscore a point that is made by many of the papers in this volume: people do not merely find themselves embodied, they actively engage in their own embodiment. The observation, while neither remarkable nor novel, cries out for precisely the sort of fine-grained, empirically grounded reflection that this volume provides.

Mema's House, Mexico City: On Transvestites, Queens, and Machos. By Annick Prieur. Chicago: University of Chicago Press, 1998. Pp. xv+293.

Miguel Díaz-Barriga
Swarthmore College

This ethnography is an engaging but uneven account of the experiences of a group of homosexual transvestites who socialize and at times reside at Mema's house. The author first met Mema, a representative for an education project for sex workers, at an international AIDS conference in 1988. The ethnography is based on a number of short trips the author made to Mexico, from 1988 to 1996, during which she conducted fieldwork at Mema's house and helped him when he was imprisoned. The author's ethnographic descriptions convey the harshness, beauty, and violence of Mema's and other key informants' everyday lives. The book, however, also reveals the need to rethink understandings of the role of theory and the politics of gay identity in framing cross-cultural research.

In the first chapter, "The Setting and the Approach," the author describes the low-income area in Mexico City where Mema lives, discusses her methodology (observations mainly at Mema's house and 18 informal interviews), and reviews the literature. The literature review provides a broad overview, focusing on gender theory and the social constructivist approach, but does not break new ground. The beginning of the second chapter, "The Everyday Life of a *Jota*" (*jota* refers to a homosexual who prefers the passive role), provides the reader with a sense of the complex and charged reality that Mema and his group face: some families uneasily accept their transvestite sons while others do not; neighbors are cordial to Mema and his friends, but some neighborhood streets are to be avoided; the threat of violence and rape is ever present. The third chapter, "Little Boys in Mother's Wardrobe: On the Origins of Homosexuality and Effeminacy," explores her key informants' narratives on how they became homosexual transvestites as well as reviews the literature on psychological and biological explanations of homosexual identity.

The fourth chapter, "Stealing Femininity: On Bodily and Symbolic Constructions," demonstrates the limitations of the fieldwork since its treatment of popular culture and discussion of image maintenance are very general. The author mentions that "the *jotas*" style themselves after soap opera stars and "follow local tastes," albeit in excess, but does not provide specific examples of how Mema and others mimic these stars (which stars?). The chapter is not helped by broad generalizations, such as the assertion that Latin-American Catholic culture is more image based than Protestant cultures (p. 145). The chapter does a better job describing how Mema and others perceive gender inequality and how, while they dress as women, they do not attempt to become women.

The fifth chapter, "*Machos and Mayates*: Masculinity and Bisexuality," explores the relationships that Prieur's informants have with *mayates* (men who have sex with other men but do not consider themselves to be

homosexual). The author attempts to link bisexuality to the "symbolic structure in which it is embedded" (p. 179). This attempt is ethnographically thin since the author mainly discusses bisexuality in terms of the literature on Mexican masculinity and gender identity. In the section on critiques of dominant notions of masculinity, for example, the author does not provide examples of how her informants resist and reproduce such identities through their everyday conversations, not to mention joking. In the next chapter, "On Love, Domination, and Penetration," the author analyzes how relations of power, often expressed in terms of penetration, condition gender and homosexual relations. In relationships between *jotas* and *mayates*, "both become victims of their own perceptions and evaluations, and every encounter is colored by schemata that give superiority to masculinity (and masculine or superiority), and inferiority to femininity (and feminine to inferiority)" (p. 254). The author ends the chapter by discussing the possibilities of more egalitarian relationships and predicts that, because of the media and other factors, the model of "the gay man" will become more prevalent. This will involve taking a step into "a model of homosexuality based on the formation of a couple with two equal persons" (p. 270). The author concludes neither by exploring the social meanings of desire nor the politics of image maintenance, but rather provides a general discussion of sociological theory's role in locating the factors that explain patterns of male homosexuality.

In this vein, the author's representation of Mema is especially problematic. On the one hand, the author treats Mema as an authority on Mexican sexual patterns, such as percentages of Mexican men who have had sex with other men. As it turns out, these generalizations are based on Mema's impressions. On the other hand, the author mentions that Mema closely read the manuscript but nowhere does she discuss Mema's comments on the book's theoretical arguments. Given the author's understandings of the identity of *jotas* and the possibilities for more egalitarian relationships, such a discussion could have enriched our understanding of how theoretical issues are framed and the politics of how identities are transformed.

Gay Politics, Urban Politics: Identity and Economics in the Urban Setting. By Robert W. Bailey. New York: Columbia University Press, 1999. Pp. xv+384.

Culture Wars and Local Politics. Edited by Elaine B. Sharp. Lawrence: University Press of Kansas, 1999. Pp. 250. \$35.00 (cloth); \$16.95 (paper).

Rhys H. Williams
Southern Illinois University, Carbondale

Both of these books are written by political scientists, primarily for political scientists, and as such illustrate both the usefulness and the redun-

dancy of much of our current academic specialization. Sociologists who are interested in either cultural politics or local/urban politics will learn some things from the political science perspectives used here. Simultaneously, sociologists are also likely to wonder whether they sometimes just are reading very familiar ideas expressed with different typologies and whether the most interesting questions that could be asked about these phenomena are in fact the ones being asked by the authors.

Both books examine the issues and politics that are loosely known as comprising the "culture wars"—in this case issues of "moral politics" such as abortion, pornography, and gay/lesbian rights. To date, these have generally been analyzed at the national level. The authors endeavor, laudably, to discover how these issues develop and are fought out in local settings. This is a favored approach of many sociologists of social movements and cultural politics; they want to know how ideology, mobilization, and political outcomes happen "on the ground" and not just in the words and actions of national elites. It is hard to think that a political issue can truly be national without involving a large number of local manifestations and confrontations. The authors here want to understand those dynamics.

Robert Bailey sets up his study by contrasting an "identity politics" perspective to the "urban economism" approach that informs most studies of urban politics. Urban economism understands the city as a site of economic development and conflict, where interest-based groups are the main players and where material conditions are the locus of political motivation, mobilization, and power. More recent identity politics is almost the antithesis of this approach, eschewing any type of class analysis for a focus on how individual and collective identity develops, how identity is translated into collective action, and how it shapes the entire political landscape. As Bailey notes, the New Social Movement theory that developed in Western Europe is the center of the identity politics analysis, while metaphors such as the "growth machine" continue to dominate urban political analysis. By developing an "identity field" theory to go with a perspective on urban governments as "regimes," Bailey hopes to bring identity and economics analyses together.

Bailey's empirical data are centered in case studies of our cities: New York, San Francisco, Philadelphia, and Birmingham, Alabama, although he also draws upon voting data for mayoral contests in Chicago, New York, and Los Angeles. The different case study cities are used as analytic narratives to illustrate different dimensions of the identity and mobilization processes involved. He portrays a continuum of the "*embeddedness of identity*" in the social organization of local governance" (p. 181, emphasis in original), running from social clubs, to service provision, to political advocacy, and varying in levels of formal organization. Thus, Bailey embeds his notion of identity fields as political terrain within an account of the organizational fields in which gay/lesbian politics develop and in which collective identity is nurtured; his account is quite resonant with

neoinstitutionalist thinking about networks, environments, and the blurring of organizational boundaries, but it is not recognized or developed as such. Rather, Bailey is more interested in variations of regime theory and understanding the impact of local context on political organizing and the affect of organized collective identity on local governance.

Bailey challenges mainstream political science with his focus on gay/lesbian identity politics. With little survey data to analyze, and unable to use the basic political socialization model of political behavior (which has difficulty accommodating developmental change), interest-group analysis and voting studies miss much of the action. Gay/lesbian residential concentration and electoral prowess are often apparent, but the extent, scope, and boundaries of "the" gay and lesbian "community" are amorphous. Here Bailey's case studies serve him well, as he can show the developmental process of sexuality as an identity marker and can examine the interactive aspect of gay/lesbian mobilization with urban governance. In sum, this is a useful and interesting book.

Elaine Sharp's collection is more tightly and thematically edited than is often the case with edited volumes. The overall focus is on local government and how local officials respond to and try to handle outbreaks of culture wars politics. To that end, Sharp presents a typology of official reactions—such as repression, evasion, responsiveness, and the like—and chapter authors examine their cases as ways of assessing the typology's adequacy. Abortion protests in South Carolina, Denver, and Topeka, gay/lesbian politics in upstate New York, and needle exchange programs in New York City are some of the cases examined; other chapters examine comparative data covering several local governments' responses to culture wars politics. The focus is usually on understanding how local government responds, and the authors often conclude that the basic typology requires some emendation. Sharp's concluding chapter summarizes and systematizes the chapter findings about local governmental issue management. She also concludes that culture wars politics deserves to be considered as a distinctive fourth arena of local politics, alongside allocational, developmental, and redistributive politics, although the actual supporting evidence in this volume is mixed at best.

Richard DeLeon's chapter deals with San Francisco's domestic partners policy; it is interesting because it reverses the dynamic investigated by the other chapters—that is, rather than examining how the national political scene affects local politics, DeLeon examines what effect San Francisco's city policy had on the national scene of domestic partners, health care, and corporate decision making. James Button, Ken Wald, and Barbara Rienzo use a large comparative data set to study gay rights politics and school curriculum issues. The comparative data put the myriad of local political battles in some kind of context and allow the authors to test statistically culturalist versus institutionalist expectations as to which types of cities develop these political controversies.

A significant drawback in both of these volumes is that they take for

granted what it is about local and urban politics that is different and possibly unique as compared to other forms and settings. What is distinct about the dynamics of local politics—about the ways in which movements mobilize, about how political authorities respond, about the life cycle of issue attention and resolution? Sharp's introductory chapter makes the standard case about why culture wars politics is different from "politics as usual," but it does not examine whether this should be any different at the local level than at the national. Several claims are made that the politics as usual of municipal settings is the "pothole politics" of service delivery, economic development, and quality of life issues. But is that true in actual city life, or is that primarily a function of the urban economism that Bailey claims dominates the field? Historical examples, usually mentioned in passing, of moral and cultural politics in places such as Cincinnati and Topeka suggest that culture wars are not that unusual at the local level. If politics is not just about material interests, but also involves moral vision and collective self-definition, then it seems obvious that this should be as relevant locally as nationally. In our current period of devolving federalism, where more decision making is going to the states, the varying local implementations of national politics and policies should be significant. How that setting is distinct is a promising area for research.

The Boundaries of Blackness: AIDS and the Breakdown of Black Politics. By Cathy J. Cohen. Chicago: University of Chicago Press, 1999. Pp. ix+394. \$18.00.

Sharon Hicks-Bartlett
Terrapin Training Strategies

Cathy J. Cohen's *Boundaries of Blackness: AIDS and the Breakdown of Black Politics* is an engaging, provocative analysis of political processes filtered through the "lens" of AIDS. Although black communities have always been diverse, political leaders have focused their efforts and mobilized around "consensus" issues that accentuated "linked fates" and the primacy of shared racial identity. To Cohen, this approach is no longer viable. Increasing diversity and widening cleavages in black communities have engendered various "cross-cutting" issues that stretch identities beyond the boundaries of race. AIDS is one example. As the number 1 killer of young black men and women in New York City, the setting of Cohen's study, one might expect fervent community mobilization. It never happened, however. To understand why black political leaders failed to accept the challenge AIDS presented, Cohen probes the roles of all the major players.

Cohen begins with the voices of those directly working and dealing with AIDS. Too often similar accounts, from both dominant and "indige-

nous" communities, mute these realities. In marginalized black communities, this silence contributes to a "secondary" marginalization that exacerbates feelings of isolation. Through these voices, Cohen unveils a landscape that augurs the complexities and multidimensionality of responses to AIDS.

Next, Cohen details a theoretical framework of marginalization. Noting its limitations, she deepens and expands her perspective on marginalization to highlight the history of power relations and oppression; the key role of institutions and organizations in marginalized communities; and the dynamic, evolving nature of strategies of marginalization (p. 25). In addition to theory, Cohen's methodology is a strength of the book. She interweaves a wealth of data including case study, roll call vote analysis, participant observations, content analysis, and interviews.

Context is important to Cohen, and the role that political context plays is of central concern. Early reports of AIDS packaged it as a gay, white male disease. Consequently, black communities and their leaders did not consider it germane. Moreover, many people viewed AIDS as God's judgment on gays and their lifestyle. Later, in black communities, AIDS became fodder for rekindling conspiracy theories. To explore the genesis of the attitudes black communities and leaders expressed, Cohen is painstaking. Early on we learn the importance of looking at AIDS responses over time. As more information became available, rigid views regarding who gets AIDS, and why, were modified. For instance, initially black political leaders vehemently and overwhelmingly rejected the needle exchange program but later amended their position. Cohen disentangles the stages of responses to AIDS and show how context, history, dominant and indigenous relations, and institutions all shape dispositions and choices.

Cohen tackles the roles of dominant and indigenous institutions and organizations. The author vividly depicts the power of these entities. The Centers for Disease Control (CDC) is one example. Early on, the CDC, collectors of indispensable data on diseases, made significant errors, which they disseminated along with other misinformation on AIDS. The CDC's institutional biases and racism blinded it to the needs of those outside the gay, white male classification they helped advance. This myopic casting of AIDS gave incentive to black leaders to avoid dealing openly with the classism, sexism, and homophobia in black communities.

In several instructive chapters, Cohen explores the role of dominant, indigenous, and alternative print and broadcast media. With the exception of the alternative press—run primarily by gays and lesbians—these outlets regurgitated the misinformation given them and perpetuated some of their own stereotypes. Indigenous media, while giving more coverage to AIDS than their mainstream counterparts, failed to provide inclusive, transformative coverage. Missing were the voices and firsthand experiences of those with HIV and AIDS hidden inside black communities.

The National Association for the Advancement of Colored People, Ur-

ban League, and Southern Christian Leadership Conference follow a pattern similar to that of the media. Cohen never insinuates that black organizations and political leaders did nothing to combat AIDS. She details their accomplishments. To Cohen, black political leaders simply did not do enough. Her analysis of roll call votes and public statements of political leaders are not only evidence of their stand on AIDS and related issues, but also of the many times they should have pushed harder. Instead of "transforming" basic myths about AIDS and mobilizing black communities, far too many black political leaders played it safe. Political leaders, rather than openly confronting the AIDS problem in black communities, allowed the fissures within communities to swell. The breakdown of black politics becomes a predictable outcome. Cohen calls for a new leadership, one that will "reflect the differential life experiences and interests among black Americans, while preserving some unified constituency available for mobilization" (p. 291).

Cohen's thought-provoking book should ignite lively debate about black political leadership and responsibility. Graduate students in political sociology, medical sociology, and methodology will find this book compelling.

Fast Lives: Women Who Use Crack Cocaine. By Claire E. Sterk. Philadelphia: Temple University Press, 1999. Pp. ix+242. \$59.50 (cloth); \$19.95 (paper).

Kathryn J. Fox
University of Vermont

Claire E. Sterk's descriptive ethnography of women from the Atlanta area who use crack cocaine illustrates the difficult and varied ways that drugs impact their lives. Sterk aims to tell the tale from the perspective of the women, thereby exploring the complexity of their relationships to crack, as well as to their significant others. The book explains some of the tensions women feel from, for example, wanting to identify as good mothers and how their drug use frustrates that identity. In addition, some of the women feel conflicted about their relationships with partners, about their illegal hustles, about the obstacles to successfully completing drug treatment, and other issues. At times, the women describe the salience of these other aspects of their lives, yet repeatedly throughout the book, their relationship to crack seems to be the most significant feature of their daily existence.

Yet, Sterk demonstrates that there is not a single, one-dimensional "type" of woman that uses crack cocaine. For example, some of the women engage in illegal hustles to support their habits while others do not; some have active sex lives with numerous partners while others do not; some try to reduce their drug use while others do not. In this respect, the book encapsulates many of the same findings that authors have previ-

ously reported about crack and other substances. For instance, Margaret Kearney, Sheigla Murphy, and Marsha Rosenbaum described the tensions involved in motherhood for women using crack cocaine ("Mothering on Crack Cocaine: A Grounded Theory Analysis" [*Social Science and Medicine* 38 (1994): 351–61]). In addition, other topics, such as the lack of subsidized drug treatment (specifically the lack of women-centered programs), have been documented by others as well. Many of the issues Sterk explores, such as the consequences of drug prohibition, are developed in a volume edited by Craig Reinerman and Harry G. Levine (*Crack in America: Demon Drugs and Social Justice* [University of California Press, 1997]). While Sterk's accounts of women's lives pulls together many of the findings from disparate sources, much of what she says has been said already.

The biggest drawback to Sterk's treatise is that it is virtually devoid of theoretical content. In a few places, the author alludes to symbolic interaction and the significance of social roles, yet she never develops an analysis of role engulfment or fully describes the process by which people become submerged into drug-using identities. For example, one chapter discusses the various ways that the women were introduced to illicit drugs, usually through boyfriends or other influential peers, but there is nothing unique about crack cocaine in this process of social learning. Thus, the author does not shed much light on the ways that some roles become more dominant while others do not, nor is it clear the extent to which these women identify primarily as crack users.

Similarly, Sterk tiptoes around the impact of race, class, and gender on these women's experiences. She mentions that these are critical variables for shaping their lives; however, she seems uncomfortable theorizing about the role they play. Sterk points out that most of the women had financial problems before they began using crack; many engaged in illegal hustles well before the onset of their drug use. There are places throughout the text where the role of poverty is brought into relief, but the process by which the experience of being shut out of the labor market, the feelings of hopelessness, are never fully mined. Although Sterk explains some of the differences in experience, for example, between male crack dealers and female ones, gender and race are both largely ignored. In the final chapter, the author reports that the women often blame sexism, racism, and classism for their problems, but this revelation is disconnected from the content in substantive chapters.

All of the chapters are arranged according to descriptive topics, such as women's introduction to drugs, AIDS, income, reproduction, and so on. The fact that there is no conceptual framework guiding the chapter organization speaks to the purely descriptive nature of the text. The concluding chapter begins to address conceptual issues by touching on how the portrayal of these women's lives reflects upon public policy about drug use. Sterk advocates a "harm reduction" approach, similar to the stance toward drugs in the Netherlands. Meanwhile, Sterk acknowledges that even if cocaine were legalized, poverty, racism, and sexism would

continue. In the end, Sterk claims that these women's "drug use can only be understood when placed in a larger societal context" (p. 215). The book does demonstrate *some* of these contextual factors, highlighting the chaos created by crack cocaine use, the lack of resources in certain communities to resist the drug trade, and the sense among residents that they have little to lose by smoking or dealing crack. Unfortunately, a concerted effort has been made to eliminate theoretical analysis whose inclusion would have made the book stronger for a sociological crowd.

Inventing Pain Medicine: From the Laboratory to the Clinic. By Isabelle Baszanger. New Brunswick, N.J.: Rutgers University Press, 1998. Pp. x+358. \$50.00 (cloth); \$22.00 (paper).

Nicholas A. Christakis
University of Chicago

Isabelle Baszanger, in *Inventing Pain Medicine*, explores the creation of the theoretical field and practical specialty of pain medicine in a way that opens up to broad and important sociological concerns, including the way professional disciplines fission and form and the way these processes interact with the creation of new knowledge and with broader socio-historical events. Scholars interested in the sociology of the professions, of knowledge, of medicine, of science, and even of social movements will be interested in this book.

This book is compendiously researched and presented, and Baszanger uses a number of qualitative techniques, including participant observation in pain clinics in the United States and France, content analysis of (American) textbooks and written materials, historical analysis, and first-hand interviewing and correspondence with some of the principals in the pain "movement." On an empirical level, Baszanger examines how and why the new conceptual and organizational forms of dealing with pain emerged at a given moment (i.e., the "invention" of pain medicine) and how and why the specific clinical interactions in this field transpire. These concerns are, in general, addressed in the first and second halves of the book, respectively. The book documents the origin and evolution of the specialty of pain medicine and the creation of pain clinics, and also the developments in biomedical theories regarding pain that both drove and resulted from these developments.

As Baszanger shows, pain is a problematic phenomenon for at least two reasons. First, it is a "private reality" that cannot be objectified given current medical technology. And second, its persistence in a patient reflects a failure of medicine to be effective. As such, it is a difficult basis for a new specialty and for everyday practice.

Baszanger shows that the challenge, when it comes to pain, is that the "physician must believe the patient" (p. 33), which is a radical proposi-

tion, as it turns out. On this point, however, the analysis in the book is a bit confused, for while the first half of the book, relying primarily on textual and historical analysis, argues that the specialty of pain medicine specifically accepted the patient's experience as the "truth," the latter half of the book, primarily based on participant observation, documents that the patient's version and interpretation of events is *not* always accepted as "real." Baszanger carefully describes the complex process by which the patient's subjective experience is mapped onto the physicians "objective" understanding. Indeed, during one clinical encounter documented by Baszanger, the physician says to a patient: "You have only your experience to go by, whereas I am familiar with many cases" (p. 149).

Nevertheless, Baszanger's work provides plentiful support for the privileging of the patient's experience as both the subjective and objective standard for the diagnosis of pain, positioning pain medicine as part of a larger modern social and medical movement that attaches great importance to the patient's subjective experience. As such, the emergence of pain medicine parallels broader developments regarding the empowerment and self-determination of patients, consumers, and citizens. Moreover, the emergence of pain medicine offers a stark counterpoint to the therapeutic hubris that was (and is) so rife in medicine, for, as Baszanger shows, the object in the treatment of chronic pain is often not to "cure" it but rather to "manage" it.

It is an extraordinary fact, as Baszanger notes, that the specialty and field of pain medicine are born of the *failure* of the prior practitioners. Since "a patient is only seen at the express request of a physician for pain that has lasted for more than six months and has proven resistant to conventional medical treatment, the point of departure is the failure of medicine at large. . . . This creates a situation quite different from other medical specialties, even if their work also includes dealing with chronic illness. . . . [Pain medicine] occupies a specific place that is linked with this idea of an initial failure" (p. 13). The book thus excellently documents how a professional failing and limitation is used as criterion to define another professional group.

Baszanger provides a compelling illustration of the ways in which the social construction of knowledge plays out in the practical world of real human beings, primarily by examining the fascinating impact of new theories of pain on both discourse and practice in the exam room. For example, she shows how the emergence of a certain theory of pain resulted, post hoc, in the legitimation of previously folk medical practices and their transformation as part of mainstream medicine. She also illustrates how definitional problems—for example, about what constitutes "chronic pain"—can both motivate the creation of a new professional form or specialty and also divide it. More broadly, however, Baszanger carefully documents the role of biomedical theory and discoveries in the creation of a "world of pain."

Baszanger misses the opportunity to connect her analysis to the work

of scholars such as Ivan Illich (who is not so triumphalist about the benefits of treating pain) or Arthur Kleinman (who situates the problem of pain in the broader problem of suffering). There is also a sense in which this work could be connected to a broader literature on the task of medicine in society. Nevertheless, this dense book is an important contribution to our understanding of the translation of scientific theory into practice, and vice versa, at the bedside of suffering patients and within the profession of medicine itself.

The Instability of Androgynous Names: The Symbolic Maintenance of Gender Boundaries¹

Stanley Lieberman, Susan Dumais, and Shyon Baumann
Harvard University

By definition, androgynous names do not serve as gender markers. Two radically different expectations about their growth are plausible: on the one hand, the rise of the feminist movement, which militates against gender distinctions, would suggest androgynous names increasing in recent decades. On the other hand, cross-cultural research indicates that first names designate gender more frequently than any other characteristic of a child or its family, suggesting a minimal increase. Examining data for all white births in Illinois in every year from 1916 through 1989 produces paradoxical results. Overall use of androgynous names is barely increasing; however, the disposition to use androgynous names has increased among parents of daughters. Analysis of the accidental ways in which androgynous names develop, their special characteristics, and their asymmetric growth patterns, leads to viewing the androgynous process as collective behavior that can be fruitfully examined through the perspective of the Schelling residential segregation model. The minimal increase in androgyny reflects a gender contamination effect that may be operating in a variety of other domains as well.

This article examines trends in the use of androgynous names during most of a century and, in turn, considers the factors affecting these developments. One can generate two radically different expectations about the use of androgynous names. Both are plausible. On the one hand, American

¹ In response to an earlier draft, Michael Hout provided very helpful suggestions that we are happy to acknowledge here, as did Tom Smith. We are grateful to Nancy Williamson for programming help, as well as her substantive suggestions. We also acknowledge with appreciation Gina Hewes for her early work with these data, the late Zvi Griliches as well as Sharon Hays for suggestions, and Catherine Kenny and Mary A. Quigley for their help in preparing the graphs and for their research assistance. Barbara Sullivan, Illinois Department of Public Health, helped us obtain the relevant data on births in Illinois. This study was supported in part by the National Science Foundation under grants SBR-92-23418 and SES-9022192. Direct all correspondence to Stanley Lieberman, Department of Sociology, Harvard University, Cambridge, Massachusetts 02138. E-mail: SL@wjh.harvard.edu

society is changing on a wide variety of dimensions so as to undercut historical dispositions that emphasize gender differences. The strength of the feminist movement, as well as the general thrust toward androgyny in other areas of taste and fashion (for example, clothing, hair, manners between the sexes, and the like), would lead one to expect a *decline* in the emphasis on gender in the naming of children accompanied by growing usage of androgynous names. Indeed, the effort to drop the inappropriate use of gender terms, an early target in the feminist movement, has been remarkably successful. This is reflected in a variety of linguistic changes: the decline in the use of personal pronouns such as his or he in inappropriate situations; the decreasing usage of marriage-linked titles such as Mrs. and Miss; the diminishing use of the term girl to refer to adult women in settings where a male of the same age is described as a man; and the replacement of gender-linked terms to ones that no longer assume the incumbents are exclusively male, as in firefighter versus fireman. Just as these and many other historic role differentiations and barriers broke down, even though at one time crossing them seemed unnatural to many, so too it is possible that this long-standing emphasis on gender marking in the use of names would not go on indefinitely—or at least not as intensely as it once did.

Moreover, we know that androgyny is by no means a recent development. Between the 13th and 15th centuries in England, the names of male saints were often given to daughters (Withycombe 1977, pp. xxxiv–xxxvi). To be sure, a modified feminine ending was used, “but it is clear that girls so named were in fact baptized and called Philip, Nicholas, Alexander, James, &c” (p. xxxv). Withycombe lists a number of other conventionally boys’ names that were used for girls in Scotland through the 17th century. We also know that disappointed parents who were hoping for a son sometimes gave their daughters boys’ names. It was in this way that *Shirley* became a female name. In the Charlotte Brontë novel, *Shirley* (published in 1849), the parents gave their baby daughter this name since they had planned to use it for the son they were expecting. According to Hanks and Hodges (1997, p. 226), this odd usage in the novel transformed the gender associated with the name. In the opposite direction, it was common for Roman Catholics in some European countries to include a female saint among the names given to sons—particularly the Virgin Mary in such forms as Maria or Marie (Weekley 1939, p. 7). Other female names, without a religious referent, were also given to sons in 18th-century England (Weekley 1939, p. 8). Since androgyny is hardly a recent innovation, it would not be surprising if an upward shift toward androgynous names is underway. In this regard, there are indeed some newly popular androgynous names in recent years: Jordan, Alexis, Riley, Adrian (Adrienne), Casey, and Dakota.

On the other hand, a cross-national comparison of 60 societies (Alford 1988, pp. 67–68) indicates that gender is the characteristic most commonly conveyed by a name—more often than any other feature of the child or the child's family. Moreover, although social class and race are major divisions within this society, the overlap between the prominent names of different classes or races is far greater than the overlap between those given to boys and girls (Liebersohn and Bell 1992). Even now, when black parents give their sons and daughters newly invented names, their construction conveys the gender of the child (Liebersohn and Mikelson 1995). Since androgynous names, by definition, are names that do not serve as gender markers, this leads to an expectation that androgyny will be relatively uncommon, or—at most—there will be only a moderate increase.

COLLECTIVE BEHAVIOR, RESIDENTIAL SEGREGATION, AND THE WORK WORLD

Our empirical analysis is guided by three important theoretical considerations: the nature of collective behavior, models of residential segregation, and gender patterns in the work world. Although these perspectives stem from efforts to analyze three seemingly different and unrelated problems, in practice there are common elements that provide a framework for understanding a number of deep gender issues—one being the typical trajectory of androgynous names in American society thus far. The maintenance of androgyny is dependent on mechanisms and collective responses similar to those affecting racial residential segregation. The ensuing collective model is relevant for a widely observed feature of gender segregation in the work world. It provides some clues to sorting out the chicken-egg causal problem for interpreting the widely observed concentration of women in less desirable jobs.

Collective Behavior

Regardless of questions about androgyny, the ever-changing popularity of first names entails matters of fashion and taste. As Coleman (1990) and others have noted, fashion and tastes are collective processes. One's choice is affected by the choices others make and, since this is the case for all others, "There is some kind of dependency among the actions; individuals are not acting independently" (Coleman 1990, p. 198). In the case of androgynous names, parents are dependent on the names that other parents are giving their children. An androgynous name that is used for daughters but no longer for sons is an oxymoron. The problem for "the boy named Sue" is not the inherent nature of the name, but rather that not enough boys have the name Sue at the same time. Had that in fact occurred, the

name would have a very different meaning for boys. If parents had an equal disposition to use androgynous names for their sons and daughters, this would increase the chances for a relatively stable level of androgyny. Due to chance factors that affect the gender makeup of the name, instability can develop such that the disposition to use the name for one sex progressively declines with each successive decline, *ad nauseam*. However, instability is even more likely when the disposition of parents toward androgyny is affected by the sex of the child. Note, this does not mean that parents must have an equal disposition to using the name for either sex. The key is that the number of parents inclined toward using the name for their daughter is equal to the number disposed toward giving the name to their son.

Applying the Schelling Model to Androgyny

We find the Schelling model of racial residential segregation (1978, chap. 4) provides a fruitful way to consider collective processes in general and androgynous naming in particular. First of all, there is a straightforward underlying model (Lieberson 1963, p. 5) of how the ethnic makeup of a residential area can radically change after a small number of a new group initially locates in a neighborhood. This involves three simple conditions: (1) their presence increases the propensity of other members to locate in the area, (2) there is a decline—or at least no increase—in the propensity of different ethnic populations to locate there, and (3) the propensity of earlier residents to move away increases—or at least does not decline. By assuming that whites and blacks differ in their disposition toward living in racially mixed areas, Schelling provides an elaborate model of how residential areas can end up far more segregated than is necessary to meet these different dispositions. The great insight provided by the Schelling model is to show how easily a racially mixed area can lose its equilibrium and become a highly segregated black area. Starting with an all-white area, as a small number of blacks move in, those whites with a low tolerance will move out, and in turn, this slightly decreases the propensity of whites located elsewhere to move in but increases (or at least does not decrease) the propensity of blacks located elsewhere to move in. With every increase in the black component, the changing composition will lead additional white residents to move out and decrease the chances that whites outside of the area will replace them. This mechanism tends to continue until the black component of the area is greater than almost any whites are willing to accept, and thereby the area becomes virtually totally black. Since there is strong evidence that white tolerance for a racially mixed area is less than what occurs among blacks, a racially mixed area has a very delicate equilibrium that is easily disrupted (see, e.g., Farley

et al. [1978, 1993], as well as an interpretation of the cause of the different racial dispositions in Bobo and Zubrinsky [1996] and Zubrinsky and Bobo [1996]).

Instead of analyzing the numbers of blacks and whites *living in each subarea* of the city, here we will ask analogous questions about the numbers of boys and girls *occupying a given name*. Just as there are some subareas where virtually all of the residents are either white or black, likewise some names are almost always given to one sex or the other. Similarly, there are names that are in varying degrees androgynous, just as there are areas of the city that are racially mixed to some extent. Likewise, we can hypothesize that parents' disposition to use a given name for their son (or daughter) will progressively decline as the female (male) percentage of users increases. As is the case for racial attitudes, the point where there is antipathy toward a name is presumably not identical for every parent. In the case of naming sons, for example, there is a continuum among parents ranging from one extreme (those who will avert from a name if they know of even one girl with the name) to the other (those who are willing to give a name to their son even if virtually all others with the name are girls). Applying the Schelling model, this would suggest the number of parents disposed to give a name to their son decreases as the name's female proportion increases. In the case of residential choices, we know that black and white tastes are not identical, with blacks being more willing to accept a higher percentage of whites in their residential area than are whites willing to accept blacks. It is reasonable to assume that an analogous process could operate when parents are naming their children. Just as blacks and whites differ in their notions of the ideal desired black-white mixture in their neighborhood and in the maximum acceptable number of the other group, a similar situation may operate for androgyny. The level of androgyny that parents either desire or will tolerate when considering a sons' name need not be the same as their disposition when naming a daughter.

A related process is also relevant here. Just as there are whites and blacks with a preference for residing in a racially mixed area, likewise there are parents who prefer an androgynous name. However, one can assume this preference is not unlimited, and at some point, the level is so one-sided that people desiring mixed areas no longer find it an attractive location. In this case, a different set of potential residents move in: namely those with a preference for little or no mixing and who therefore find this lopsided distribution of their own race an attractive feature. Similarly, increasing numbers of parents preferring an androgynous name for their offspring will find the name unattractive as its usage becomes progressively too lopsided to meet their goal. As this process moves on, the name's potential source of attraction is to parents preferring a name with minimal

androgyny, and where the dominant gender having the name is of the same sex as their child. Leslie, for example, at one point might have been appealing to parents desiring an androgynous name, but its usage at present is so overwhelmingly for daughters that it is unlikely to meet a taste for a high level of androgyny.

How can a model based on a frequently occurring event—changes in residence—be applicable to a phenomenon (personal names) in which people rarely change from their initial condition? Residential location has a fluidity that is largely absent from first names. There is no inherent limitation to individuals relocating from one area to another—indeed, people rarely live in the same place throughout their lives. By contrast, although names are also changeable, most people seem to keep the same one throughout their life span. Accordingly, compared to the influence of neighborhood shifts on residential movement, changing tastes in names are far less likely to generate an analogous shift.

In fact, the two processes are strikingly similar once one recognizes that the appropriate application of the segregation model to tastes is somewhat counterintuitive. The response to changes in the gender usage of one's given name is *not* the appropriate application of the Schelling model—indeed the resemblance is only superficially the same. The appropriate analogue is the response of those giving names to children (in other words, parents). The issue is whether a model dealing with how adult residents respond to changes in an area's racial composition is applicable to how parents respond to changes in a name's gender composition.

A change in the gender composition of a name alters the inclination of parents to give the name to their newborn son or daughter. This entails a mechanism that is exactly the same as the way residents respond to racial changes in a neighborhood. Just as it is probably rare for parents to change the existing names of their children, likewise it is unlikely that one's residential history will be altered in response to current developments in a neighborhood. Rather, the changing racial composition of a neighborhood (or changing gender composition of a name) respectively will impact on ensuing decisions with respect to residence (or naming).

A second question stems from the straightforward information mechanism operating in the Schelling model whereby residents are able to gauge changes in the racial composition of a neighborhood. Through simple observation, most residents will form an opinion about both the area's existing racial makeup as well as recent changes. Can an application of the Schelling model also assume an information mechanism about gender changes in a name's usage that is as straightforward as the way it occurs for racial composition? Androgyny encompasses a scope that is far greater than one can observe directly, involving far more infants and small children than those living in one's neighborhood.

This "difference" is deceptive on several counts. First of all, it is probably difficult for residents to gauge precisely a neighborhood's racial composition. They may be fairly confident about the extremes—virtually all white, virtually all black, fairly equal numbers of both—but finer gradations are less accurate. Likewise, conclusions about recent changes are probably very much influenced by contacts on one's own block and knowledge about specific friends or acquaintances in the area. Although information about a name's popularity in general—to say nothing about its relative popularity for boys and girls—entails events occurring elsewhere than in one's neighborhood, such information is partially available. One will encounter children's names in a variety of settings: the names given to children by friends and acquaintances living elsewhere, casual encounters such as parents addressing their children in public places such as zoos, beaches, shopping centers, toy stores, and the like. Also, the usage of a name in the media (as a character in a popular television program or movie, or in a newspaper account, or as the name of a movie star or other popular figure) will impact on one's impression of how the name is used.

There is an even more important feature—one that is basically analogous to estimating an area's racial composition. Whether it be the racial composition of an area *or* the gender usage of a name, estimates are based on incomplete knowledge drawn from the relevant cases that are encountered. Residents draw conclusions based on changes in specific housing units, neighbors, and people that they know. However, they also draw conclusions from encounters on the street, in local stores, schools, and the like. Likewise, impressions about the gender usage of a given name are based on casual encounters in addition to children and adults that one knows.² In either case, then, impressions are derived from incomplete information. Although one's sources are unlikely to be random—whether it be the racial makeup of a neighborhood or the gender usage of a name—at least they *are* in part a function of the actual distribution of the underlying population. If the name in question is given to a child—whether it be son or daughter—then presumably the parents have encountered the name at least once. As the popularity of the name increases differentially for sons and daughters, then the probability is that: (1) parents will encounter the name in greater frequency and (2) they will encounter the name in greater frequency for one sex than the other. In this regard, the issue is not the completeness of the information for all children with a given name, but the actual information encountered.

² This is especially the case for names that are not part of the standard repertoire of names, say John, Mary, Michael, Jennifer, etc. Below, we will see that it is just these newer nonstandard names that are particularly likely to become androgynous.

Finally, the Schelling model assumes that the racial makeup of a neighborhood is important—in varying degrees of intensity—to most residents or would-be residents. In the same fashion, it appears that the naming of a child is not a casual matter for most parents. After all, an infant's name is a reflection of the parents' taste and disposition. Even if we ignore parental concerns about the name's possible impact on the child in the years ahead, in that sense the child's name is a reflection on parents no differently than their choices in clothing, automobile, furniture, and the like. Accordingly, it is reasonable to consider the application and relevance of the Schelling model of racial segregation to a domain that on first glance appears so remote from its initial concern.

If this approach is valid, two important predictions follow: (1) it is difficult for a name to remain androgynous, and (2) the movement away from androgynous names will not be the same for boys and girls. The instability of androgyny is due to the following feedback process: as the proportion of girls, for example, with a name increases, this will decrease the likelihood that sons will be given the name and may—if anything—increase the disposition to give the name to daughters. In turn, this further increases the female proportion, and that means an even larger segment of parents will find the name unattractive for newborn sons, and so on. Although the model is applicable to either sex, there is every reason to assume for the moment that the disposition toward androgynous names will—for the same parents—be affected by whether they are naming a son or a daughter. As a consequence, when androgynous names unravel into a single-sex name, the direction of that outcome will be affected by gender differences in the number of children of the opposite sex that will move the name toward exclusive use for boys or for girls.

Gender Segregation

Our analysis of androgyny, and its possible linkage with the Schelling model, may also be relevant for understanding gender segregation in other settings where a variety of causes also operate. There are numerous factors, for example, that lead women and men to differ greatly in their occupational distribution. These include the desirability of the occupation with respect to income and other economic considerations, the prestige of the work, the opportunities for advancement, disposition of employers, resistance from current workers, stereotypes about the gender suitability for the work, and the ability of men to select more attractive occupations (see, e.g., Cohn 1985; Jacobs 1989; Reskin and Roos 1990; Reskin and Padavic 1994). Certainly, gender composition itself may play a symbolic role in drawing males away from an occupation simply because the presence of a large number of women is unattractive to them. However, given the

wide variety of these other considerations, it is difficult to sort this out in the work world. A noteworthy exception is Williams (1989), whose study of women in the Marine Corps and men who are nurses leads her to observe "that men in both the Marine Corps and the nursing profession do make greater efforts than women to distinguish their roles from those performed by the opposite sex" (p. 10). We cannot resolve this issue here but do note that many of the organizational influences on employment are absent in the naming process. For example, there are no employers, no unions, no income differences that are attached to a name. The naming process, particularly with respect to androgyny, provides a more direct opportunity to observe the possible influence of gender composition in a situation that is relatively insulated from the institutional factors operating in the job market. Under those circumstances, is there evidence that at some point the presence of women acts as a negative force driving males away?

TRENDS

The first step is to determine both the level and trends in androgyny for a long span of years. To our knowledge, the information currently available on androgyny is either anecdotal or based on less than adequate measurements of androgyny. The pioneering effort of Barry and Harper (1982) compares the names recommended for both boys and girls in three naming books published between 1933 and 1946 with three others published between 1969 and 1979. Since their conclusions are not supported by a comparison with actual naming practices in these periods and since there are no quantitative data on the total number of children given these names, the results are of only modest value for gauging trends.³ Barry and Harper (1993) improve on this early effort with a second study, which compares actual naming choices in Pennsylvania in 1960 and 1990. However, their conclusions are unwarranted in many instances because they overinterpret very small numbers of cases. At the most extreme, Jan is classified as moving from a girls' name in 1960 to a boys' name in 1990 based on the fact that three of the five children given the name in 1990 are boys (Barry and Harper 1993, table 2). Moreover, their analysis is restricted to two points in time that are 30 years apart. As will be evident later in the article, looking at names year by year provides the only way of examining the implications of the Schelling model and, in turn, obtaining a deeper understanding of the processes influencing androgyny.

³ Moreover, of the three books for the earlier period, one is a publication from England and another is a second edition that was published after the early period it was meant to measure (Barry and Harper 1982, p. 16).

The analysis here is based on Illinois Department of Public Health birth certificate data on the name and gender of every child born in the state between 1916 and 1989 (as well as 1995).⁴ To our knowledge, this is the first instance where a time series of this magnitude is reported for androgyny trends anywhere in the United States. Because of the enormous increase in newly invented names among blacks, from the 1960s onward (Lieberson and Mikelson 1995), we elected to concentrate exclusively on whites at this point. However, no exclusion is made on the basis of ethnic origin; accordingly, Hispanics in Illinois are included if white is the race reported on the birth certificate.

In contrast with studies based on changes in the names that popular naming books claim are androgynous, we can determine the actual frequency of androgyny in every year separately for each sex and thereby measure trends through the years.⁵ Moreover, we also employ a variety

⁴ The initial analysis, based on data for 1916–89, led us to wonder if developments in the late 1980s was the beginning of a new upsurge in androgyny. By special arrangement, the Department of Public Health graciously provided us with the relevant data for 1995, the most recent year then available. This is the only exception to our year-to-year analysis and is indicated with a separate line on the relevant graphs.

⁵ Even these data are less than ideal. Careful analysis of the Illinois Department of Public Health data tapes suggests that a small number of errors occur in the gender coding of births in the initial years, which declines steadily over time. Miscoding of gender will lead to an artificial boost in the level of androgyny. Consider the impact of such errors in the most extreme case where each name is given exclusively to either boys or girls, with no overlap. Each error in the entry of the sex code would incorrectly create an androgynous name. Such errors, at least at a very low level, are plausible for any data set with massive numbers of entries. They would be especially plausible for earlier in this century, when the recording of birth certificate data almost certainly involved the use of punch cards as an intermediate step between the original birth certificates and the tapes that were eventually generated later by the Department of Public Health. They are much less likely in recent decades where there is less chance for human error in the transmission of the raw birth certificate data into computer-readable form. Since the Department of Public Health has no record of the early procedures, it was necessary to estimate the degree of gender miscoding. A list of prominent boys' and girls' names (20 each) was generated. Making the arbitrary assumption that these names are unlikely to be used for the opposite sex, we ask how many cases are reported as such and, further, whether the usage declines in a manner to support our hypothesis. The evidence is, indeed, that errors were greater during the earlier period, with the decline in such errors more or less paralleling the decline in the androgyny index. At the beginning of the period, the tapes classify about 4 per 1,000 of children with such names as Mary, Jane, and Irene as boys; likewise, about 4 per 1,000 of children with such names as William, Robert, Paul, and James are reported as being girls. Applying the same set of 20 boys' and 20 girls' names, the estimated level of misreporting drops steadily, first reaching 2 per 1,000 in 1925, 1 per 1,000 in 1930, and is persistently less than 1 per 1,000 since the mid-1940s (often considerably less). These estimated rates based on the names specified above were then applied to all reported births in each year in an adjustment procedure, which assumed they were representative of the larger population. The impact of the adjustments are virtually nil after the first few decades, but they have a bearing on the

of sources to analyze other characteristics of androgyny. Such an undertaking, involving almost 11 million white births between 1916 and 1989 (and 1995), is only possible because the Illinois Department of Public Health had entered the information from every birth certificate into a machine readable format and then made them available to the investigators for analysis.

It is easy enough to develop an operating notion of what is meant by androgynous first names; these are names commonly given to both boys and girls such that it is not a clear marker of the bearer's gender. Straight-forward as this may be, it is another matter to develop suitable measures of androgyny. First, how equal does the gender distribution have to be in order for a name to be considered androgynous? Presumably, most readers would have no difficulty with a name given to 50 boys and 50 girls born in a given year. Since such an exact level of equality is unlikely in practice, at what point should the deviation from exact equality lead us to apply the androgyny label—60:40 (yes), but 99:1 (arguably no)? Second, how many children need to be given a name before it can be labeled as such? Should a name given to two boys and two girls born in a given year be considered an androgynous name? One can make a number of differing arguments on these matters, but inevitably it involves a certain level of arbitrariness that in turn can affect the results. The difficulty is resolvable if we draw a distinction between the level of androgyny in the entire population of newborns in a given year and, on the other hand, prominent androgynous names (a characteristic that we operationalize later). Finally, following a common practice in naming studies, all forms of a name are pooled together—as long as the pronunciation cannot be distinguished.⁶ Hence, androgyny refers to the level of gender overlap of a given name as pronounced—regardless of spelling. Later in the article, we will address the role of spelling in androgyny.

Turning to the first matter, we want to know the level of overlap between the names given to boys and girls. One extreme, the absence of any semblance of androgyny, would occur if there is no overlap between the names given to boys born in a given period and those given to girls, that

earlier period. Accordingly, the graphs in figs. 1 and 2 are based on adjusted Illinois data. It has no bearing on the curve shown in figure 1 (given the scaling), but it does mean that the surprising results in fig. 2 are not an artifact of data errors. Indeed, without these adjustments, the reported androgyny index would be actually slightly higher in the earlier period. (The low level of estimated error in the later decades, during a period when androgyny goes up again, is harmonious with our assumption that these are largely errors in gender coding for these major names, rather than androgyny.)

⁶ In practice, some of these are pronounced differently for females and males (as is discussed below).

is, total segregation. No segregation—the other extreme—occurs if every name is given equally to boys and girls. At this extreme, for example, the sex ratio of children named Michael or John would be the same as the sex ratio of Jennifer or Mary. A priori we know that neither extreme exists, but what is unknown until now is the actual level of gender overlap and the changes over a span of roughly 80 years. The Index of Androgyny is derived from the P^* index (a measure described by Bell [1954] and later modified to its current form by Lieberman [1980, pp. 253–57]) that is widely used to gauge spatial isolation among racial and ethnic groups (see, e.g., Massey and Denton 1993) as well as gender isolation among occupational groups (see, e.g., Jacobs 1986, 1989). The Index of Androgyny measures the overlap between the names given to white boys and girls born in each year between 1916 and 1989, as well as 1995. It is the equivalent of determining, for the name of every individual girl born in a given year, the boys' proportion of children with that name. The index gives the average of all these proportions. This value can range from 0 (if girls all have names that boys never receive) to approximately 0.51 (if the boys' proportion of each name is identical to the male percentage of all births).⁷ The index will generate virtually identical results if we focus on the names given to boys in the same year and instead determine the percentage with each name who are girls.⁸ The analogous summary measure will range from 0 (if all boys have names that are never given to girls) to 0.49 (which, given the sex ratio at birth, would be the girls' component of every name if it were identical to their percentage of all births). The slight difference between the two maximums reflects the sex ratio at birth, and we address this by averaging the two index values, which then leads to a range from 0 (no androgyny) to 0.50 (for complete androgyny).

Two striking features of androgyny during this century are evident in the moving averages shown in figure 1.⁹ First, androgyny is uncommon

⁷ Based on a sex ratio of 105 at birth.

⁸ The equation for the Index of Androgyny is

$$P_b^* = \sum_{i=1}^n \left(\frac{g_i}{G} \right) \left(\frac{b_i}{t_i} \right),$$

where G is the total number of girls born in a given year, g_i is the number of girls with a given name, b_i is the number of boys with the same given name, t_i is the total number of children with the name. The Index of Androgyny for boys is obtained by substituting the genders in the same formula.

⁹ Moving averages are a widely used procedure in statistics, economics, and the natural sciences for describing the trends over time by minimizing (or "smoothing out") minor yearly oscillations. The five-year form for the Index of Androgyny is computed by summing the first five consecutive years and then dividing by five. This value is given for the midyear of the five-year period. In turn, the first year is dropped, and a sixth year is added, and the process is repeated again, ad nauseam. By necessity, no moving

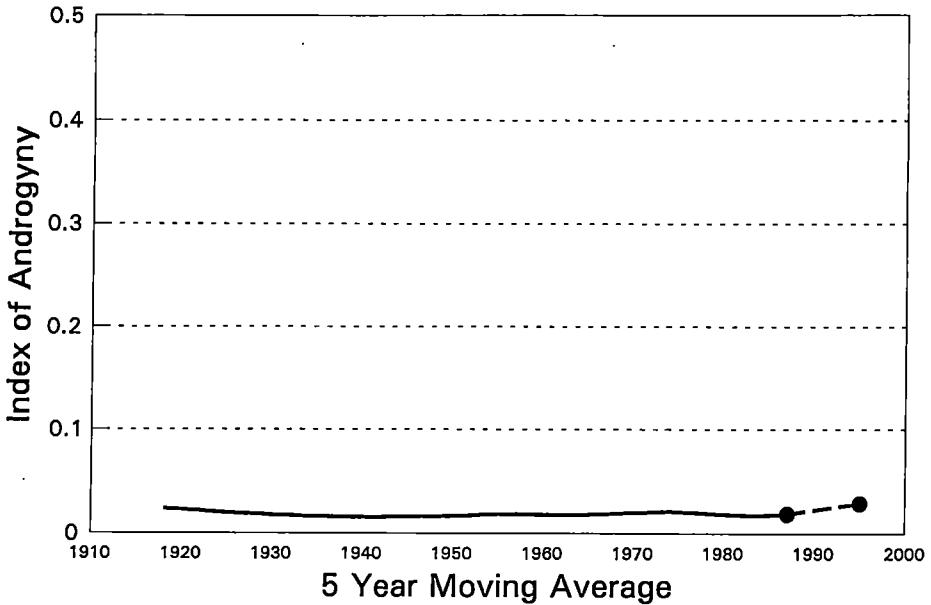


FIG. 1.—Level of androgyny, Illinois, 1916–1989, 1995

throughout the 80-year span, at least when viewed in the context of the range of possible values. The index is about 0.03, at its highest during this span, which means that for the average girl (or boy), less than 3% of the children with her (his) name are of opposite sex. Looking at the index values that would occur if there was total gender segregation of names (0.00) or no gender segregation at all (0.50), we see that the actual level is less than 6% above the value for total segregation. In other words, 97% of the average child's name is given to children of the same gender. The second important feature of figure 1 is the absence of any obvious trend through the years. From the perspective of the total range of possible values, changes are minor and barely visible when scaled across the total range of possible values. Androgyny is relatively uncommon—even at present.

Variations over time do appear if we focus on the actual values within this narrow range. However, the curve in figure 2 is an odd one. Androgyny, at the beginning of the period, is about as high as it has been until very recently (fig. 2). A persistent decline occurs for a number of decades, extending to the Second World War. This is a surprising result. Tom

average value is possible for either the first two years or for the last two years of the available raw data.

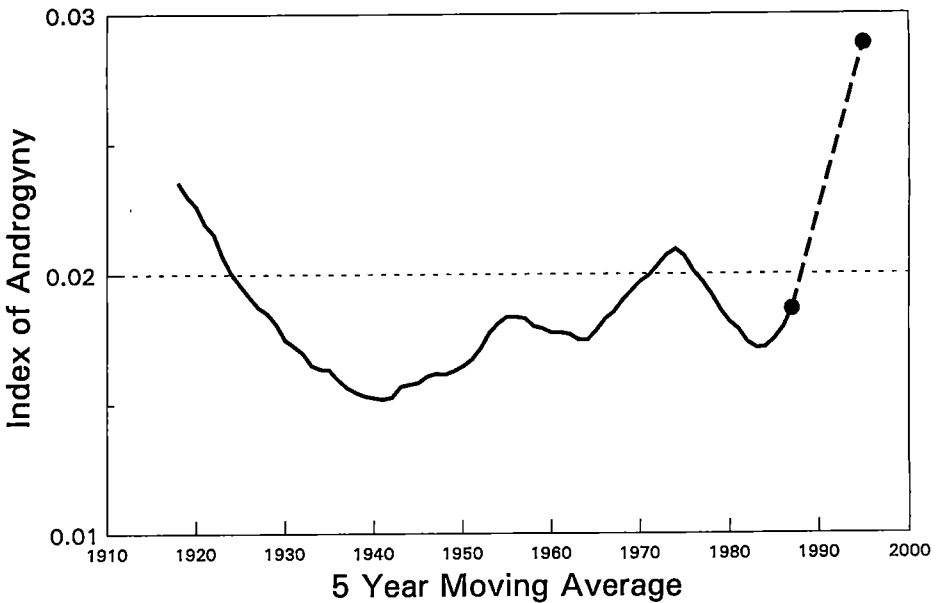


FIG. 2.—Enlarged view of changes in androgyny, Illinois, 1916–1989, 1995

Smith (in a personal communication) suggests that European immigrants may not have been fully aware of the usual gender connection of some names in the United States and, in turn, unintentionally gave some American-born children a name not usually associated with the child's gender. Moreover, in some cases, the cognate of the name used in different European sources may not have been associated with the same gender, as is the case for Carol, which is used for Charles in Polish and Rumanian (see Stewart 1979, p. 74). Such errors would lead to a higher androgyny index that would decline as immigration declined and the earlier immigrants became acquainted with American practices. We are unable to evaluate this intriguing thesis because the computerized Illinois data during this early period do not report birthplace of the parent. However, this speculation is harmonious with other results reported below about how androgyny often develops.

There are three waves of androgyny since World War II.¹⁰ From the late 1940s to the mid-1950s, there is a rise that then declines through the mid-1960s. A substantially larger increase in the androgyny index then

¹⁰ It may be inappropriate to attach too much substantive meaning to these fluctuations, given the narrow range of changes. Bear in mind, however, that each yearly index value is based on a very large number of cases, to wit, all of the white births in that year.

occurs for about 10 years, beginning in the mid-1960s. This turns downward, back to its original level. Compared to other values in this narrow range, the 1995 index indicates a *relatively* large increase in androgyny. Indeed, the Index of Androgyny for 1995 is higher than that occurring in any other year—regardless of whether one considers either the specific yearly index or the values for the moving average. Obviously it is impossible to determine if the 1995 index will mark the beginning of a long-term change in androgyny or will be part of a cycle similar to the earlier one. Ultimately, this is an empirical question. However, we will be able to provide some clues to the social processes that have been generating these cycles.

An Additional Indicator of Androgyny

Before engaging in a more detailed evaluation of possible causes and a theoretical interpretation of this “nonevent,” it is prudent to be confident that the conclusion is not an artifact of the method. Accordingly, we briefly consider names that are unambiguously androgynous—that is, that are simultaneously used for boys and girls in a relatively balanced way. By “balanced,” we mean where each gender comprises at least one-third (and hence, no more than two-thirds) of the children bearing the name. Following this approach, we still reach the same conclusions. First, the proportion of children given such androgynous names tends to follow the temporal pattern reported above in figure 2; second, the proportion of all children given such names is—at best—modest. In the earliest available year, 1916, 2.1% of infants were given such a name. This moves downward to 1.5% of children born in 1953 (about midway in the span), 1.1% in 1989, and then turns up to 1.6% in 1995.¹¹ Again, androgyny is relatively infrequent and exhibits no long-term upward movement.

SPECIFIC ANDROGYNOUS NAMES

Why does androgyny remain so infrequent? This seems like a naive question. Usually, our interest is in why some event occurs, rather than why it does not. Certainly the outcome is consistent with the long-standing and widespread practice of giving children gender-linked names. However, this is not a very satisfactory explanation. The absence of androgyny cannot be used as an indicator of widespread antipathy toward gender-neutral names—that would be a glaring case of circular reasoning in

¹¹ There are 208 names that meet this standard in the earliest year, 85 in the middle of the span, and 92 most recently. It is misleading to simply focus on the *number* of names given relatively equally to boys and girls, since many of these names are numerically unimportant.

which the effect is used as a measure of the cause. Moreover, even if there is not a major shift in naming patterns, it is puzzling that so little change occurs in response to changes in gender relations during recent decades. The summary measures are, by their very nature, too abstract to provide us with an understanding of the intermediate mechanisms and processes that generate this outcome. We know the net outcome, but we need to look at specific names themselves.

To answer these questions, we need to trace the life course of specific androgynous names that are at one time or another at least minimally popular for both boys and girls. We want to know where androgynous names come from and then what happens to them such that the level of androgyny is so modest and the shifts over time relatively minimal. In particular, we need to consider whether there are any distinctive characteristics of androgynous names themselves that would help account for the low levels of androgyny. Only then, after taking these characteristics into account, can we determine how the collective mechanisms discussed at the outset help us understand the life of an androgynous name.

The criterion for inclusion as an androgynous name is straightforward. For at least one of the years between 1916 and 1989 (as well as in 1995), the name is among the 200 names most frequently given to newborn boys *and*, likewise, girls (although not necessarily the same year for each sex). By this standard, there are 45 androgynous names that appear in Illinois during the span under consideration (table 1).

These 45 androgynous names are rarely even modestly popular for both girls and boys in the same year (that is, ranking among the top 100 names given to newborn boys *and* girls). In Illinois, there is only a slight overlap in the prominent names given to white boys and girls throughout the period under study. To be sure, there is at least one name found on the list of 100 most common names given to both newborn girls and boys in Illinois data for every year except two (1985 and 1988), but it is only a slight overlap—it varies between one and three shared names in any year. Put another way, of the 100 most popular names for newborn girls or boys in a given year, typically between 97 and 99 are not found in the top 100 for the other gender—keep in mind this is after pooling all possible spelling forms. Altogether, there are 13 different names that are one of the 100 most common boys' and girls' names in the same year. About half are in the top 100 for both sexes for relatively long spans. These are: Frances/Francis (1916–53, 1955), Marion (1916–26, 1928, 1930, 1932), Jean/Gene (1927–44, 1946–47, 1950), Carrie/Cary (1958–65), Terri/Terry (1947–72), Jamie (1970–84, 1986–87), and Jordan (1995–?).¹² These are the only an-

¹² For the most recent year examined, 1995, Jordan is rising rapidly in usage for both sexes, albeit favored more for sons than daughters (respectively the thirty-ninth and

TABLE 1
45 LEADING ANDROGYNOUS NAMES, ILLINOIS,
1916–1989, 1995

"Traditional" Names	"New" Names
Alexis	Adrian/Adrienne
Billy/Billie	Angel
Carroll/Carol	Bobby/Bobbie
Dale	Cary/Carrie
Dana	Casey
Francis/Frances	Corey/Cori
Gale/Gail	Dakota
Gene/Jean	Devin
Jackie	Jaime/Jamie
Jerry/Jeri	Jan
Jesse/Jessie	Jordan
Jody/Jodi	Kelly
Joe/Jo	La Verne
Kim	Loren/Lauren
Lee	Marlin/Marlyn
Leslie	Riley
Lynn	Sean/Shawn
Marion	Shannon
Merle	Stacy
Rene/Renee	Taylor
Robin	Tracy
Sidney/Sydney	
Terry/Terri	
Tony/Toni	

NOTE.—Male and female forms shown when the most common spelling is different. Traditional and new names defined by their presence or absence in Stewart (1979).

drogynous names that are given to at least 0.2% of both boys and girls born in the same year. These results are not terribly different from the pattern in New York State, from 1973 through 1985, where there was not one overlap between the 100 leading names given to boys and girls (Liebersohn and Mikelson 1995, p. 933).¹³

Bear in mind that it is fairly common for these 45 androgynous names

fifty-first most popular name). At this point, obviously it is impossible to determine its long-term stability as a name widely in use for both sexes. The other names are Casey, Kelly, Kim, Leslie, Shannon, and Tracy.

¹³ The criteria were more restrictive since it was done for each specific year and spelling variations were not taken into account. In Illinois, as explained earlier, all spelling variants generating the same sounds are pooled (e.g., Frances and Francis).

to reach the top-100 level of usage; 28 at some time or another reach the top 100 for sons and 27 reach the top 100 for daughters. Even when they do reach this level for both, it is rarely at the same time. For our entire set of 45 names, there is no correlation between the highest rank a name reaches for girls and the highest rank the name reaches for boys (Spearman's $\rho = -.17$; $P = .13$). The high point of the names' popularity for one sex is simply unrelated to how popular the name will be for the other.

CHARACTERISTICS OF ANDROGYNY

The specific characteristics of these 45 androgynous names help us understand how androgynous names typically develop. Most of the names have at least one characteristic that is favorable to androgyny. Traditional sounds associated only with names given to one or the other sex tend to be absent; the names frequently originate through separate pathways such that each gender has a distinctive claim for the name; many are spelled differently for each sex, and hence a separate gender marker can remain even though the pronunciation is identical; and an exceptional number of these are nontraditional names and therefore escape a historically strong gender connection. Yet, their modest later developments are all the more puzzling since their characteristics tend to be gender neutral.

Unanchored Names

Androgyny is especially prominent among what we will call *unanchored names*, that is, names with characteristics that are not markedly linked to one gender, and hence are less likely to have features that clash with any implicit expectations parents may have about what seems an appropriate name for either their daughter or their son. Accordingly, they can be more readily used for both sexes.

New names.—Androgyny is relatively more common among newer and less traditional names—names that have less of an established gender association. Of the list of 45 major androgynous names (table 1), close to half (21) are not found in *American Given Names* (Stewart 1979), a volume that we can use as a source of traditional names used through the late 1970s. By way of comparison, respectively 91% and 86% of the 100 most frequently used boys and girls names are included in the aforementioned book. Compared with the leading names given to boys or girls, these newer names (as indexed by their absence in Stewart) are more prone to become androgynous.

Sounds.—There are some sounds that are strongly associated with names given to one of the sexes. Of particular interest, some of the different endings found in names are implicitly understood by the population

TABLE 2
THREE MOST COMMON ENDINGS AMONG 45 LEADING
ANDROGYNOUS NAMES, ILLINOIS, 1916–1989, 1995

	45 LEADING ANDROGYNOUS NAMES	TOP 100 NAMES FOR EACH SEX	
		Boy	Girl
% ending with:			
\bar{e}	42	15	24
n	29	18	24
l	11	11	7

to be girls' or boys' names—witness the way newly invented names follow the same gender patterns as those exhibited among very common names (Lieberson and Mikelson 1995). Androgyny rarely occurs among names with a strong gender-specific association. One such association is the use of *a* endings. In Illinois, among the 100 most commonly used names for girls throughout the period between 1916 and 1989, 25 end in *a*, whereas this occurs only for Joshua among the top 100 boys' names. This is perhaps the most obvious example, but there are other linguistic differences between the endings of boys' and girls' names (see Lieberson and Mikelson 1995). The striking feature about the 45 androgynous names is that their endings are rarely of the type that is gender linked. Three different sounds account for the endings among 80% of the major androgynous names. These are \bar{e} , n , and $(ə)l$ (as in, respectively, Leslie, Gene/Jean, and Gale/Gail). All of these are common to both boys' and girls' names (table 2). For this feature, too, androgyny tends to occur among names unanchored in a strong gender-linked matrix.

Parallel Pathways

Often androgyny develops through parallel pathways, that is, through the separate usage for girls and boys of what is more-or-less the "same" name. There is either a gender marker such that the name's application for girls and boys is slightly different in its form or in the original names from which they evolve. In either case, the androgyny develops from distinctively different origins for the male and female usages. Indeed, although we are unable to evaluate this question with the available data, some cases of androgyny may well develop simply when two different and independent pathways merge and thereby generate an androgynous name. About 40% of the 45 names become androgynous through some sort of parallel

TABLE 3
LEADING ANDROGYNOUS NAMES WITH AN \bar{e}
ENDING, ILLINOIS, 1916-1989, 1995

DIFFERENT ENDINGS*		SAME ENDINGS	
Girl	Boy	Girl	Boy
Billie	Billy	Kelly	Kelly
Bobbie	Bobby	Stacy	Stacy
Carrie	Cary	Tracy	Tracy
Jamie	Jaime	Lee	Lee
Jessie	Jesse	Casey	Casey
Cori	Corey	Riley	Riley
Jerri	Jerry	Sydney	Sidney
Jodi	Jody	Jackie	Jackie
Terri	Terry	Leslie	Leslie
Toni	Tony		

* For these names, "i" or "ie" is the predominate spelling for girls, but not for boys.

process, such that an intentional copying of the other gender's usage is not necessarily at work. In other words, the "same" name becomes a boys' and girls' name through separate paths, rather than simply the expansion or imitation from one gender usage to include the other as well.

Spelling.—In the United States, as well as other English-speaking nations, the spelling of names operates in a laissez-faire environment. Consequently, many popular names are spelled in a variety of ways even if used primarily for only one sex. Often one form is by far the most common one, although there are a variety of less important spelling variations that are nevertheless pronounced in the same way as the dominant form. Androgyny often occurs when there is a difference in the predominant spelling of the male and female versions of the same name. In nearly half of the 45 leading androgynous names (21), the most common way of spelling the name is different for boys and girls. These differences can occur in any part of the name, but spelling variations that mark off girls from boys are especially pronounced in the endings. It is common for both girls' and boys' names ending with the popular \bar{e} -sound to end with the letter *y*.¹⁴ However, among androgynous names ending in an \bar{e} -sound, there appears to be a strong tendency for an *ie* or *i* spelling to occur among girls and a *y* to be found among boys, with other substitutions among less common spelling forms as well toward *ie* or *i* for girls (see table 3). These minor

¹⁴ Among the 200 most commonly used nonandrogynous names between 1916 and 1989 in Illinois, there are 23 boys' and 23 girls' names that end in *y* or *ey*.

spelling variations allow androgynous names to retain an element of gender specificity.

Diminutives.—Ignoring androgyny for the moment, the reader should note that many first names develop diminutive forms that serve as an informal alternative to the formal name (i.e., the name on the birth certificate). For example, Jim is used for James, or Peggy for Margaret. In many instances, these diminutives evolve into standard names themselves and become part of the existing inventory of formal names (i.e., they occur on birth certificates). For example, Lisa, Betty, Eliza, Liz, Beth, and Elsie were once diminutives for Elizabeth but are now names in their own right. In many cases, both a boys' name and a girls' name separately develop diminutives that are similar—even if the original names are not. In other cases, the diminutive for one name is identical to an existing formal name for the other sex. In both instances, an unintended consequence of nicknames becoming formal first names is that sometimes they overlap with an independent development for a name given to children of the other sex, or overlap with a name already in use for the other sex. (The origins of the specific names discussed in this section on diminutives, and the following one on other parallel paths, draw from information in at least one of the following sources: Dunkling and Gosling [1983], Hanks and Hodges [1997], Stewart [1979], Withycombe [1977].)¹⁵

This is relevant to our discussion here; many of the androgynous names (table 1) were initially diminutive forms of names that themselves were distinctive and different for the two genders. Through parallel paths, they evolve from names that are themselves not androgynous. For example, Bobby and Bobbie were initially used for Robert and Roberta respectively. Other examples are: the male usage of Gene stems from Eugene, thereby meshing with the historical usage of Jean for females; Jackie comes separately from Jacqueline and Jack; J(G)erry stems not only from boys' names (Gerald and Gerard), but also from Geraldine; Jesse for boys has a biblical origin, whereas Jessie for girls originates as a diminutive for Jessica; Jody is originally a diminutive for Judith and also for Joseph and Jude; Jo(e) is a diminutive derived from Joanna, Josephine and other female names, but also from Joseph; Kim is from both the male and female usages of Kimberly as well as Kimball for boys; Lynn, a name used for males, became a female name after -lyn was used as a suffix for female names and eventually became a name itself. In the case of Marion/Marian, the latter female usage is long-standing, but the initial male usage of Marion derives from the surname of Revolutionary War hero, Frances

¹⁵ The accounts are not always consistent and are sometimes based on limited information, but one or more sources are harmonious with the processes indicated here.

Marion, 'the Swamp Fox.' Terri(y) separately served as a nickname for Theresa and Terence before developing into a formal first name for either sex; Toni(y) stems from Antonia and Antoinette for females, and Anthony for males; the male usage of Stacy derives from its being a nickname for Eustace, the female usage stems from its initial usage as diminutive for Anastasia.

Ethnic differences.—Androgyny can develop when the "same" name occurs as male in some nations (or languages) and female in other countries or languages. This does not usually generate androgyny, because in most societies there is little ambiguity about the gender linked to the name. In the United States, however, which receives large numbers of immigrants from divergent sources, the very same name may become androgynous because of these differences in usage. The androgynous nature of Carol illustrates European influences coupled with the evolution of the female form from nonethnic sources within the United States. Carol, spelled in a variety of ways, is a version of Charles occurring among various European ethnic groups.¹⁶ On the other hand, the usage of Carol for females probably stems from its being a shortened form of Caroline (Withycombe 1977, p. 58). As Carol becomes an exceptionally popular name for daughters, the clash between the two usages leads to the decline of the male application. A vestige of the European influence is still found; eight of the 17 boys in Illinois given the name between 1985 and 1988 are of Polish origin.

Several other androgynous names stem from ethnic differences in the gender of a name that is spelled identically—or virtually so—and where the pronunciation is, in varying degrees, different for the two sexes. Two of these, Angel and Jaime, reflect differences between Spanish-speaking groups and the remainder of the white population. Angel is a boys' name in Spanish (pronounced as if it were spelled "anhel"), whereas it is used for girls elsewhere. This difference persists. More than 95% of the 317 Hispanic children given this name are boys. (As above, ethnic tabulations are for Illinois births between 1985 and 1988.) When non-Hispanic parents use Angel, the concentration is almost as great, but in the opposite direction (90% of the 334 children are girls). Clearly, the androgynous usage of Angel reflects ethnic difference in the use of the name; this is coupled with different pronunciations, such that male and female versions remain distinct—on the oral level.

Two different naming paths for James also intersect in the United States. Jaime is the Spanish and Portuguese form of James. On the other

¹⁶ Carol in Romanian; Karel in Estonian, Danish, and Bohemian; K(C)arel in Dutch; Karol in Polish and Slovakian; and Carroll among the Irish (Yonge 1884, p. 386; Dunkling and Gosling 1983, p. 41–42).

hand there is Jamie, similarly spelled, but initially used as a diminutive for the English James, which then developed into a girl's name (Hanks and Hodges 1997, p. 124). In turn, a variant of Jamie develops—Jaimie. As is the case for Angel, androgyny is facilitated because the Spanish Jaime is pronounced as if it were “haheemeh”—radically different than the customary pronunciation in English of an initial *j*. Not only does androgyny develop, but the customary differences associated with the spelling forms have begun to dissolve in the non-Hispanic population. Almost all Hispanic children named Jaime are boys (only 3 exceptions among 240 children). Although Hispanics rarely use any other forms, when they do, it is usually for daughters (31 of 44 children). Other whites—for whom the name is less deeply rooted in tradition—are less apt to use the Spanish spelling, and when they do it is *mainly for daughters (185) rather than sons (35)*. Indeed, non-Hispanics favor these forms for both sons and daughters.

Jan is another example of a name that becomes androgynous because two separate ethno-linguistic practices simultaneously operate in the United States; in this case, the customary use of Jan among a number of European immigrant groups clashes with developments within the English-speaking world. Jan is a form of John found in a various parts of Europe.¹⁷ This usage for sons was commonly pronounced as if the name was spelled “yan.” By contrast, the usage of Jan for daughters stems from a completely different source, namely as a spinoff from Janet, once an extremely popular female name. Again, we see differences in pronunciation helping to facilitate—for a brief period—the name's popularity for both sons and daughters. Except, of course, the sons probably have parents from European locations where Jan appears, and the daughters are likely to have parents from other sources or—if not—who are removed from the immigrant generation.

Finally, note in all of these cases that the usual outcome occurs even when an ethnic difference in the gender associated with a name is also accompanied by distinctive ways of pronouncing the name. An increase in the name's usage for one sex still leads to a decline in its popularity for the other.

Other parallel pathways.—In some cases, other independent paths generate androgyny. Robin, for example, was a boys' first name—derived from serving as diminutive for Robert. Its initial use for girls appears to reflect an earlier tendency to adopt bird names for naming daughters. The mass media also affect androgyny. Tracy, for example, was originally a

¹⁷ Yonge (1884, p. 45) includes Welsh, Breton, Dutch, Belgium (presumably meaning Flemish), and Poland.

male name. It initially was used for females through the diminutive pathway, to wit, as a nickname for Teresa. In reviewing this list, one should not put too much emphasis on the direction of the shift (that is, the adoption of a female nickname into a name corresponding with a male name is more frequent than the opposite transition). Gender differences in the disposition toward androgyny are a vital consideration, but we have to analyze them through a different approach. This is because diminutives (and other nonstandard names) are more likely to be given as formal names for daughters than sons, even if androgyny is not an issue (Lieberman and Bell 1992, p. 548). However, the key point is that androgyny is especially likely to occur among names that are *unanchored* or reached androgyny through separate *parallel paths*.

APPLYING THE SCHELLING MODEL TO ANDROGYNY

Why is it difficult for androgyny to flourish? The answer certainly does not lie in the characteristics of the names themselves. If anything, androgynous names have characteristics that tend to minimize an association with just one sex. Nor does the answer lie in parental disposition to give these 45 names to daughters. Their usage increases during the period: from about 3.5% of newborn girls at the outset, it reaches a peak of almost 11% for girls born in the mid-1970s and tapers down to 7% in 1995 (this is still roughly twice the level at the beginning). On the other hand, there is only a modest increase for sons: from about 3% at the outset to a peak in 1995 of 4.7%. This only tells us that a difference existed; it does not tell us *why*.¹⁸ (These figures may seem paradoxical, given the minimal shifts in the androgyny index. This occurs, as is shown below, because names that were once androgynous in practice often end up as names used primarily for only one gender—more often daughters. As a consequence, an overall measure of gender segregation, based on the actual composition of names, will not necessarily go up even with an increase in the percentage of newborns given such names.) From the perspective of collective behavior, we recognize that individual choices affect the choices that others make, and these, in turn, will affect additional choices, and so forth. Moreover, when subpopulations differ in the levels acceptable to them,

¹⁸ The occurrence of new androgynous names during the span is complex since many of the earlier names are later more-or-less out of favor and, on the other hand, many of the more recently popular androgynous names were not in use earlier. The key point is that parents' usage of the same set of names for their daughters increases far more than for sons.

they will respond differently to the choices made by others. Under the conditions of the Schelling model, the final outcome will not necessarily be what would occur if a mechanism existed to minimize the total gap between each person's ideal and what each person gets.

The Schelling model is normally applied to different groups (as in the case of black-white residential segregation). Here, in the case of androgynous names, we are not discussing *different parents*, but rather *parental differences* in their disposition toward giving an androgynous name to a newborn daughter as opposed to a newborn son. There is no reason to assume that these dispositions for sons and daughters are identical. In other words, the ideals regarding androgyny that parents will hold when naming a daughter will not necessarily be the same as the ideals when naming a son. As is true for segregation, under these circumstances the Schelling model would predict that androgyny would often be an unstable state since parents would refrain from giving a given name to children of one sex when they would still find it attractive for the other gender. This, in turn, would lead to the same feedback process as occurs in the residential segregation of the races. Suppose, for example, parents will tolerate a larger percentage of opposite-sex naming occurrences in giving a name to their daughters than they will tolerate for sons. *Ceteris paribus*, more often (but not always) names will drift toward becoming increasingly a name given largely to daughters. This is because, as the usage for sons decreases, additional parents will then find the new gender composition unattractive for sons, and this in turn will make the name unattractive for other parents of sons and so forth.

As is the case for the residential segregation of groups, where one considers the racial makeup of a given neighborhood in terms of changes in each group's percentage of all residents, we are also interested here in the male/female composition of a name and how it changes. However, unlike a neighborhood where it is convenient to make the simplified assumption that the absolute number of residents over the short span is limited by a constant housing stock, this is obviously not the case for a given name. There is no reason to assume that the number of children with a given name will itself be constant—the number may fluctuate over time in ways that are far less likely for the number of residents in a neighborhood. Accordingly, we need to consider how the stability of an androgynous name is influenced by both the percentages and numbers of boys and girls. In either case, the evidence points toward the same conclusion. At some point, one gender reaches a certain level of usage that then becomes a barrier to expansion by the other group, the latter then declines, and an androgynous name (in terms of its early usage) progressively becomes a name that is favored for only one sex.

Percentages

Tracing each of these names over time, we find that names newly used for both boys and girls tend to be "unstable," in the sense that there is rarely a parallel upward movement. Most often, one gender's usage increasingly dominates, and the other's usage declines, or flattens out, or—at most—goes up at a modest and much slower clip. Of the 31 names starting from a common position of minimal usage for either gender (and eventually given to at least 0.2% of the children), in only 4 cases is the 0.2 level reached for both boys and girls—in the remainder, it is only for boys (10 names) or only for girls (17 names). Figure 3 provides examples of two names that become predominantly girls' names (Casey and Dana) and two that are primarily for boys (Angel and Sean). They are, of course, especially interesting cases with a pronounced breakout for one sex and not the other. However, in one form or another, the vast majority of names (27 of 31) start at a minimal-level usage, become androgynous by the standards described earlier for the 45 names, and then expand differentially such that gender usage widens over time. Although these initially neutral names increase more often for daughters than for sons, the same barrier operates. It is uncommon for an androgynous name to exceed the 0.2% level of usage for both sons and daughters. After one sex reaches this level (which means being given to 1 child per 500 births among either daughters or sons), it is rare for its usage to also reach that level for the other sex. In brief, androgyny per se is largely an unstable state: the names display a certain pullback in the usage for one sex if the other one begins to take off. Either the male or the female usage of the name may gain a certain level of acceptance, but rarely do they both achieve a similar level.

Numbers

An analysis of these androgynous names in terms of their absolute numerical values leads to the same conclusion. The numerically dominant gender associated with an androgynous name is rarely replaced by the other sex so long as the first gender's usage is at even a modest level. This holds for both traditional and recently invented names. It holds, as well, for names that simultaneously increase from minimal usage for both sexes. Only 4 of the relevant androgynous names are given to at least 100 sons and 100 daughters in the same year.¹⁹ In the remaining 32 cases, the gender first reaching the 100 level is the one that continues to ascend.²⁰ In all of

¹⁹ Carrie/Cary, Kim, Leslie, and Terri/Terry are the only exceptions.

²⁰ Patterns for the remaining nine names are not applicable: seven were already declining by 1916 (Frances, Dale, Gene, Marion, La Verne, Lee, and Merle), and therefore

these cases, the name first reaching the 100 level is never challenged by the other sex for at least the entire period when it is at this numerical threshold. This is illustrated by the patterns shown for three of the names in figure 4 (Corey, Kelly, and Robin) and the apparent trajectory for Taylor. Incidentally, in the four exceptions, it is female usage of the name that passes the male level at a time when it is used for 100+ sons. This is harmonious with a conclusion that parents have lower antipathy toward giving a name to their daughter that is also used for a fairly large number of boys. A lower tolerance exists for giving a sizable number of sons a name that is already in use for at least 100 daughters each year.

Entertainment and Public Attention

For some names, the movement away from androgyny reflects the name's usage by either a new star or as the name of a character in a major motion picture. In effect, the usual expansion of information about a name's gender connection(s) is accelerated by the national attention given to the name's usage in a particular gender context. This can serve to speed up its movement toward popularity for one sex (or even reverse a downward trend). The Schelling model is still applicable—it is just that mass attention on the name's gender usage can speed up the process leading to the decline in androgyny. This is analogous to “block busting” in a neighborhood. Typically realtors use various devices to frighten whites into panic selling at unrealistic prices—thereby pocketing a handsome profit when reselling to blacks. This includes spreading both true information about the movement of a black family into an area, as well as creating the impression that a rapid shift is occurring. The point is that information spreads far more rapidly than it would otherwise and thereby sets in motion a faster progression through the stages of change—which incidentally reduces the chances of a multiracial equilibrium being reached.

Kim (shown in fig. 5) is especially interesting for this reason. It is one of the four exceptional names with more than 100 sons and daughters simultaneously given the name (for three years). Through 1953, Kim is favored more for sons than daughters (153 and 90, respectively, in that year). The graph appears to be moving in the usual pattern: the growth in the name's usage for daughters slowing down as the number given to sons passes the 100 mark. However, a surge in daughters named Kim starts in 1954—the year when a new actress, with the stage name of Kim Novak, first appears. Novak becomes immensely popular almost immediately and is the highest ranked box office attraction among actresses by

no conclusion is possible for them; and two other names are not relevant because neither gender reached the 100 level for usage in any year (Marlin/Marlyn and Riley).

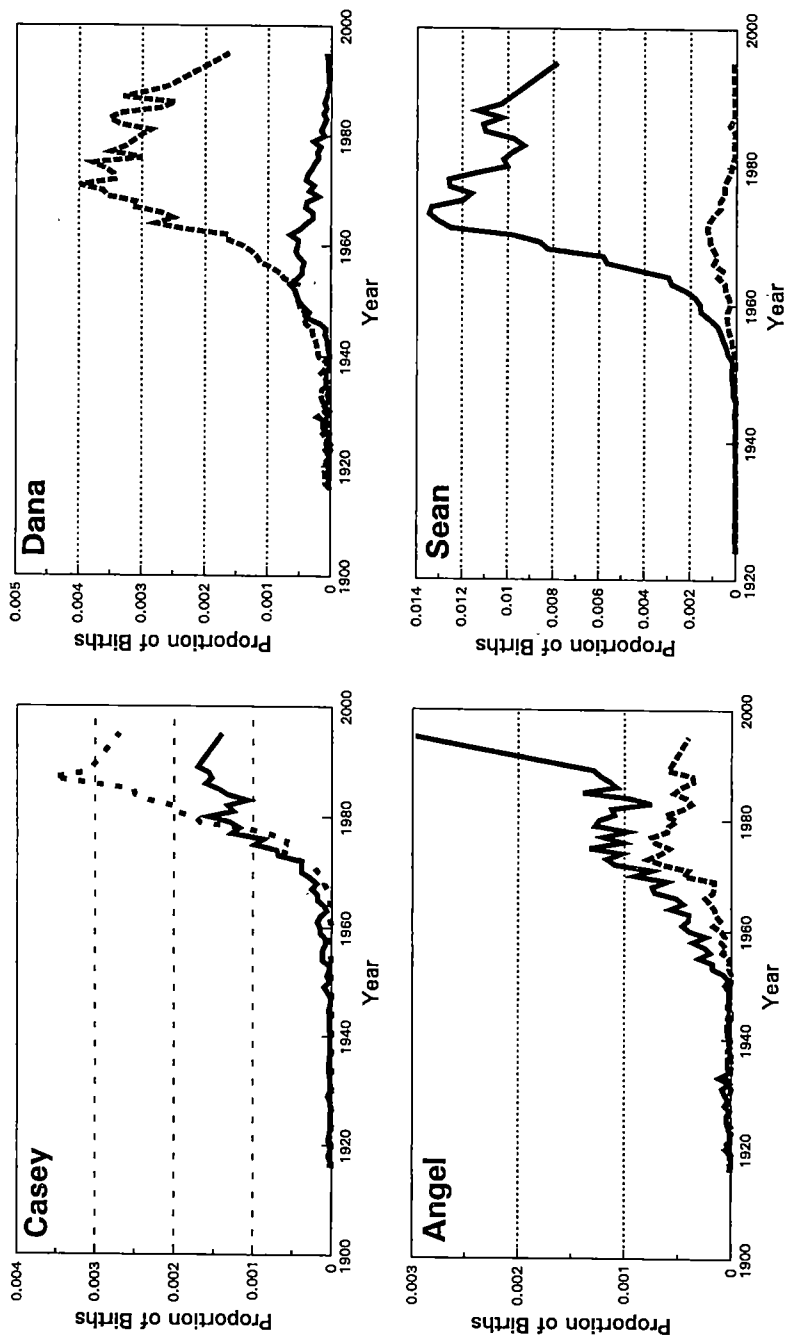


FIG. 3.—0.2% tipping point, Illinois data (boys = solid line; girls = broken line)

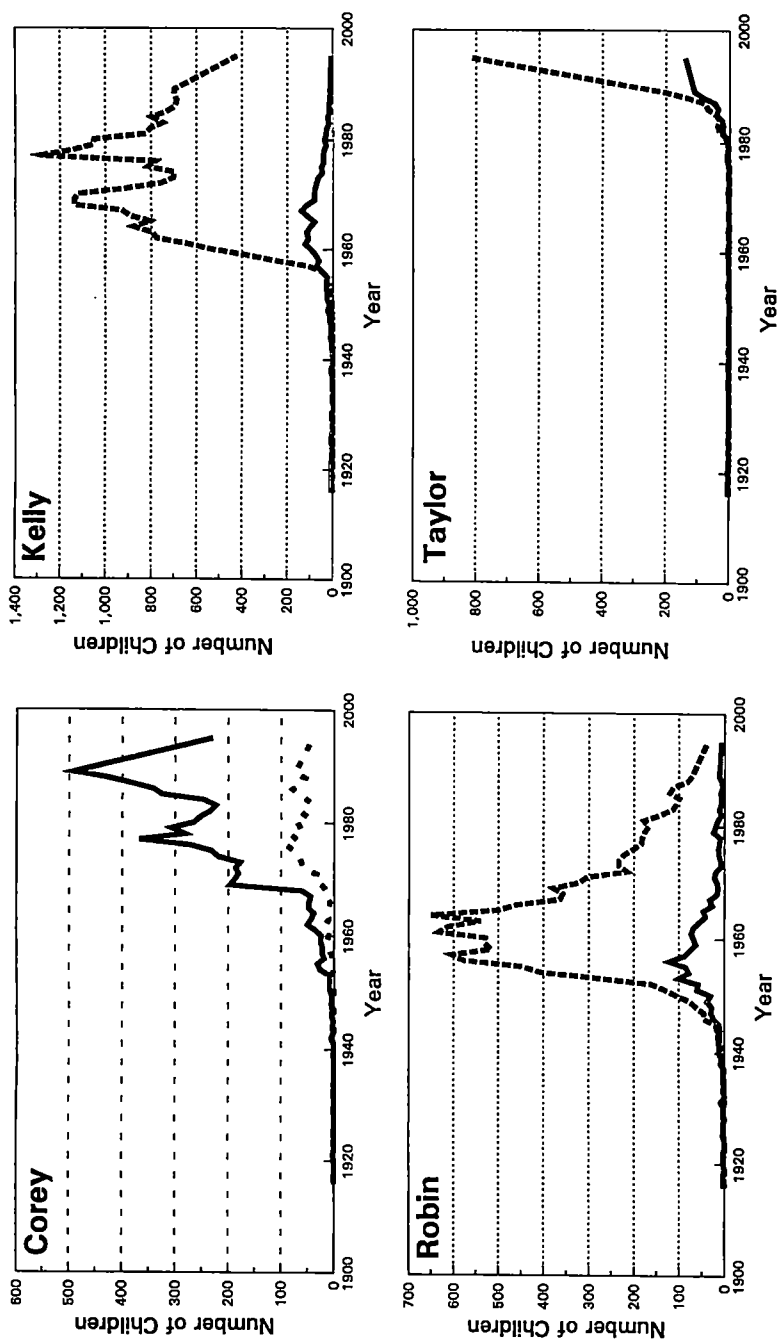


FIG. 4.—A numerical obstacle, Illinois data (boys = solid line; girls = broken line)

1956 (Peary 1978, p. 181). These developments totally disrupt the existing pattern. The number of boys given the name in 1955 is barely greater than in the previous year and then starts a downward plummet—going to 111 in 1956 and to 45 by 1958. During the same span, the name gains rapidly for daughters. The precise timing with the actress' debut and almost instant popularity is hardly coincidental.

The Tracy graph in figure 5 provides another example where androgyny is disrupted by the media—in this case Tracy Lord is the name of the lead female character in a popular 1956 movie "High Society." (Moreover, the character is played by a popular actress of that era, Grace Kelly.) Prior to this period, as the graph indicates, Tracy was a minimally popular name. This movie kicks off a persistent growth for daughters; at its numerical peak in 1970, the name is given to almost 1,500 daughters. By contrast, the increase for boys promptly slows down and reverses after a few years. As is the case for Kim, there is a correspondence with the media event.

Too much can be made about the Hollywood influence on androgyny. Witness the minimal impact of Cary Grant on the name's usage for boys (fig. 5). Although Grant's successful career starts in the early 1930s, the name's growth is fairly parallel for both sexes during the next 20 years—albeit slightly more popular for sons. Starting in the late 1950s (a period in which he is still very popular), the name loses favor for sons and becomes an important name given to daughters. Lauren is another example, where the decline of androgyny cannot be attributed to the appearance of a movie star. Although Lauren Bacall becomes a major movie star when she debuts in 1944, there is only a modest impact, and in the following year, 43 daughters are given this name. There are only modest changes for 30 years, with rather minimal gender differences in the name's usage. It is only 30 years after Bacall's debut that Lauren becomes increasingly popular for daughters.

INFLUENCE OF CHILD'S GENDER ON THE TASTE FOR ANDROGyny

A key assumption in the above analysis is that parents find androgynous names more appealing for their daughters than for their sons. It is one matter to observe differences in the numbers of daughters and sons given androgynous names, but there is a danger of circular reasoning in using this fact as a measure of the relative appeal of these names. It is another matter to determine if there is *independent* evidence to support the conclusion that these patterns reflect differences in the taste for androgyny.

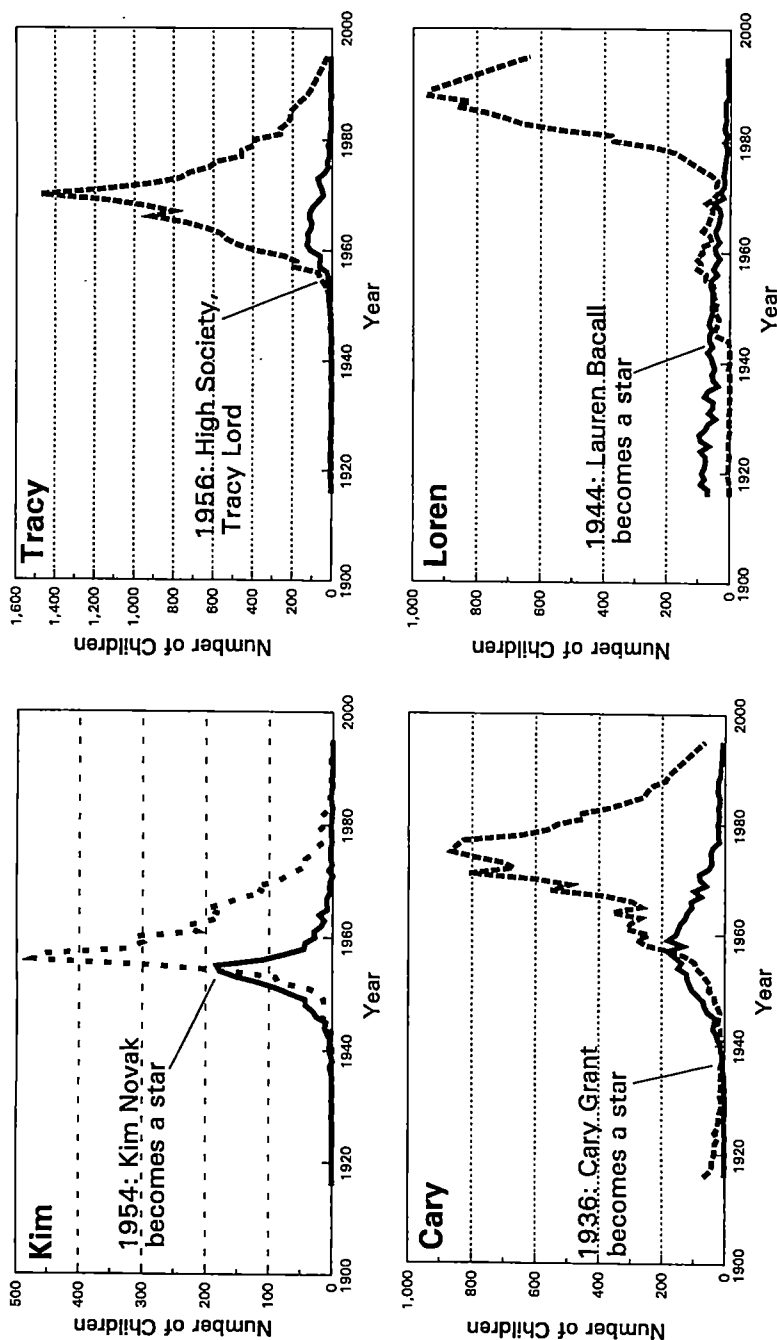


FIG. 5.—The influence of movies, Illinois data (boys = solid line; girls = broken line)

Imagery

The imagery associated with androgynous names helps us understand why they are more popular for girls than boys. Buchanan and Bruning (1971) studied 1,380 Ohio State undergraduates (both women and men) to determine the characteristics they associated with more than 1,000 names. Of the basic set of 45 androgynous names analyzed here, 23 are found among the names they used for males and 26 are in the female listings.²¹ The study asks subjects to rate names on three characteristics. For two of these traits, both males and females with androgynous names are especially attractive, whether categorized as a female or male name. On the *active-passive* dimension, the vast majority of androgynous names are rated above the median name of each sex.²² The subjects also *like* the androgynous names more than the median name for each sex.²³ For the third attribute, the rating of each name on a *masculine-feminine* scale, masculinity for the majority of the male androgynous names scores below the average for all males; and likewise the majority of female androgynous names are below the median femininity level. The majority of the 23 male names are seen as less masculine than the median male name (16 and 17, respectively, for male and female respondents). Similarly, respondents score most of the androgynous names as less feminine when used for a female (of the 26 names, 22 and 18 are below median when rated by, respectively, male or female respondents). It is as if these names, because females and males share them, have their femininity reduced when given to girls and their masculinity reduced when used for boys. One can speculate that this reduction is particularly likely to make the names unattractive for sons. Undoubtedly, for many parents, the girls' names also become less attractive because of the lower level of femininity associated with them, but we speculate there are a number of parents for whom a muted emphasis on femininity would make the name a more attractive choice for their daughters. For example, Lieberman and Bell (1992, pp. 538-41) find highly educated parents tend to favor names for daughters that have

²¹ The data were never published in their entirety, but James L. Bruning generously provided the senior author with a copy of the entire set.

²² Of the 23 androgynous names given to males, male raters score 16 above the median activeness level for all male names (15 by female raters). Of the 26 androgynous names listed as names given to females, 21 are rated above the median by female subjects (and 22 by males).

²³ For the *like/dislike* dimension, between 20 and 19 of the 23 male androgynous names are scored above the median (male and female raters, respectively); for the 26 female names, 23 and 19, respectively, are scored above the median for all female names.

connotations of strength and other attributes that were traditionally associated with males.²⁴

Supporting this interpretation is the fact that the level of masculinity attributed to each of these androgynous names affects their usage among white boys born in Illinois in 1970. The Spearman rho, between number of sons given each name and the masculinity attributed to it in the Ohio State study, is 0.56 ($P = .003$). By contrast, the usage of these androgynous names for daughters is less influenced by the femininity attributed to them (Spearman $\rho = .29$; $P = .08$). To be sure, androgynous names are seen as less masculine (when used for males) and less feminine (when given to females) than are names given exclusively to either sons or daughters. However, a low level of masculinity has a far greater negative impact on their usage for sons than do perceptions of low femininity affect the use of androgynous names for daughters. (Bear in mind these are the images associated with each name—this need not be an accurate characterization of the personality traits associated with androgynous first names.)²⁵

Also, earlier we noticed that 21 of the 45 androgynous names are spelled differently for girls and boys. Carrie and Frances, for example, are usually given to daughters; Cary and Francis to sons. There are also 24 androgynous names where the predominant spelling is the same for both sons and daughters, for example, Dana and Kim. Whether spelled identically or not, the names usually reach a higher level of popularity for daughters than sons. Indeed, if anything, names with only one dominant spelling reach a higher level of usage for daughters than do names where the spelling forms serve to differentiate gender of the child (the median rank is 54 and 60, respectively), but the difference is minor.²⁶ By contrast, androgynous names with different spelling forms reach a higher level of popularity for boys than do names where there is no differentiation by spelling (median ranks are 72 and 93, respectively). This result, similar to evidence obtained from the stereotypes, is consistent with the assumption that fewer parents find androgynous names appealing for sons and, therefore, a spelling differentiation has a certain appeal because it minimizes gender confusion. Obviously, more study needs to be done on this, but

²⁴ Data limitations of changes in the imagery and androgyny over time prevent us from examining an alternative interpretation, to wit, names with such images are more likely to be given to both sexes to begin with.

²⁵ Indeed Rickel and Anderson (1981) report no association between subjects' score on the Bem masculinity/femininity scale and their having an androgynous first name.

²⁶ Where rank 1 is given to the most widely used name in a given year, 2 is second most, and so forth.

at least there is plausible evidence pointing to reasons why this might occur.²⁷

Influence of Gender and Androgyny on the Life of Names

A central assumption is that androgyny is evaluated differently, depending on whether parents are naming a daughter or a son. We have seen that parents of daughters also respond to the number (or percentage) of boys with the name, but they are slower to retreat from using it. As a consequence, androgynous names end up as a predominantly female name more often than as a predominantly male name. Although the results so far are consistent with the Schelling model, the longevity of names can provide direct evidence of the differential disposition toward names. We compare the life expectancy of boys' and girls' nonandrogynous names with the expectancy of androgynous names for each sex. By the phrase "life expectancy," we have in mind applying to names a very specific meaning used by demographers in the analysis of mortality: the average number of years lived by a set of names from their birth until they all die. We consider the "birth" of an androgynous name to be the year when for the first time the name ranks among the 200 most frequently given names to both girls and boys. (A name that is in the top 200 for only girls—or only boys—is not yet counted as an androgynous name.) Starting with the year of birth, for each sex separately we determine the number of years the name is in the top 200. "Death" is defined as the year when the name no longer ranks among the top 200 given to a specific gender. In other words, the life span of the name can be radically different for each sex if it remains among the top 200 for one sex long after it drops out for the other. For nonandrogynous names, we select other names for each sex that also had the same ranking when the androgynous name first reached the top 200.²⁸ Because of a truncation problem—some of the names have not yet left the top 200 rank at the end of the period under study (resulting from either their exceptional longevity or their recent birth)—we use the standard estimating procedure for determining the survival function with censored data.

²⁷ The entire distribution of dispositions and acceptable levels for parents is yet to be investigated when it comes to giving a name to one's daughter as contrasted with one's son.

²⁸ Of the 45 androgynous names, 43 are included for daughters and 42 for sons. Exceptions are Dakota and Riley since they were "born" in 1995 and hence it is not possible to apply hazard models to estimate their eventual lives. Also, Sydney never reached the top 200 simultaneously (its usage in the top 200 for girls first occurs after it had dropped out of the top 200 for boys). Accordingly it was included as an androgynous name for daughters but not for sons (since parents giving the name to their daughters were presumably aware of the earlier "traditional" usage of the name).

Among nonandrogynous names in our matching sample, boys' names have a much longer average life than do those given to girls, respectively, 47 versus 30 years before disappearing from the top 200 ($P < .001$, single-tailed test). This is consistent with the general observation that names popular for boys have a slower turnover than do girls' names—presumably reflecting the greater emphasis on fashion in selecting girls' names and the disposition of parents to prefer the names of grandfathers and fathers and other more traditional names for their sons (see, e.g., Rossi 1965; Lieberman and Bell 1992). The results are radically different for androgynous names: their average years of life jumps to 54 for girls, compared with 30 for nonandrogynous girls' names ($P = .001$). By contrast, if anything, androgynous boys' names live shorter lives than do their nonandrogynous counterparts, declining from 47 to 37 years of life ($P = .14$).

The influence of androgyny on the life of a name (girls increasing and boys declining) is described graphically in figure 6. The uppermost graph compares the survival functions of nonandrogynous male and female names. The survival ratio is the proportion of names that are still in the top 200 after the indicated number of years since birth. A higher level of survival for males is displayed through much of the span.²⁹ The gender-specific survival curves for androgynous names, in the bottom panel of figure 6, show a slower decline for females than occurs among the nonandrogynous. The bottom line, in terms of the question at hand, is the difference between boys and girls in the usage of an androgynous name. On the average, we see a somewhat lower survival ratio for these names among boys in the first few years after the name is born (that is, reaches the top 200 for both sexes). In turn, during the following years, there is a considerably steeper decline for boys ($P = .10$, single-tailed test).³⁰

These results confirm the earlier assumption that a very different disposition exists toward giving an androgynous name to sons and daughters. On average, androgynous names tend to quickly lose their appeal for sons, whereas parents persist in giving their daughters the same androgynous name for a much longer span. For many of the names that have a *history* of androgyny, the net result is that they become increasingly favored by parents for their daughters, but not their sons.

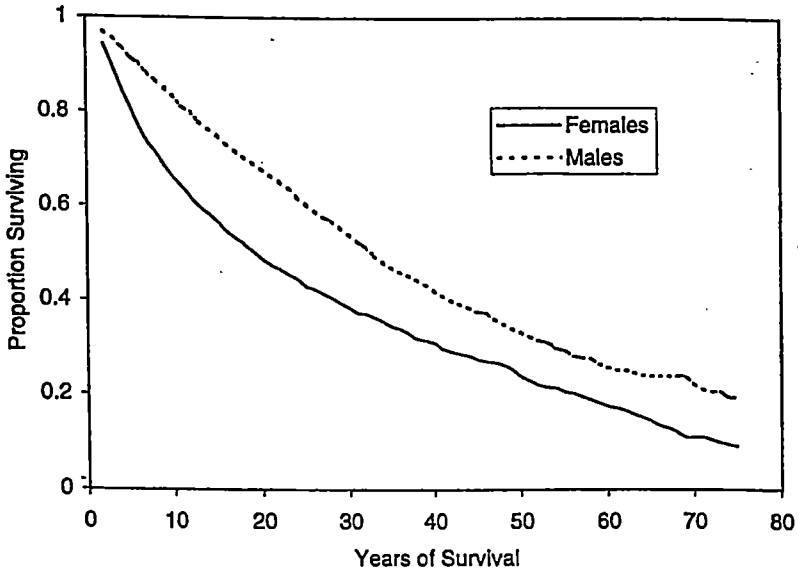
COMMENT

This study starts with a straightforward question about whether there is increasing use of androgynous names in recent decades. There is a reasonable basis for expecting a substantial increase, but on the other hand,

²⁹ Eventually all will die, and the survival ratios will be identical for both sexes.

³⁰ As indicated above, bear in mind that this comparison between male and female androgynous names is based on a small number of cases.

Non-Androgynous Names Survival Rates



Androgynous Names Survival Rates

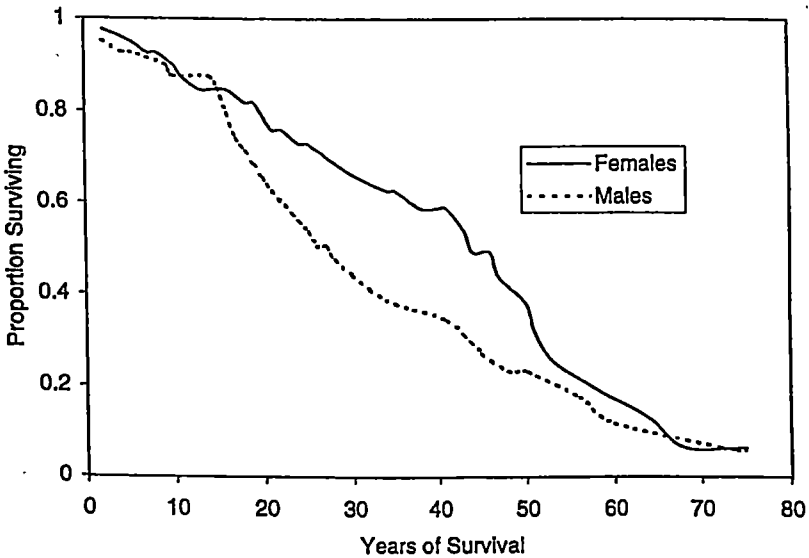


FIG. 6.—Influence of androgyny and gender on the survival of names, Illinois, 1916–1989, 1995.

there are good reasons for expecting no increase. Overall, we have seen relatively small changes in the frequency of androgyny. Paradoxically, although there is an increase in the percentage of daughters who are given "androgynous" names, there has been a very minimal increase in the level of androgyny. How can this be? Unlike issues of racial residential segregation or issues of gender segregation in the workplace, here the interests of one group are not pitted against another in a straightforward manner. Rather, naming entails contradictory dispositions *within* the same group. Among the daughters born in 1995 to white mothers who are at least college graduates, 8% were given one of the 45 androgynous names listed in table 1. On the other hand, these same androgynous names were given to 3% of their sons—a figure that is actually lower than that for mothers with lesser levels of education. When parents are more inclined to use an androgynous name for their daughters than their sons, this favors an unstable outcome. A name that is initially androgynous will not remain androgynous indefinitely if it is given to daughters in expanding (or even stable) numbers at the same time that parents begin to avoid the name for their sons. Moreover, although the disposition toward androgyny is stronger for daughters than sons, it is not without bounds for daughters. We encounter a number of cases where the names for sons expanded more rapidly, and in turn this led to a decline in usage for daughters. All of this generates a situation that makes it very difficult for androgyny to flourish. For both sons and daughters to receive the same name at the same time, it is necessary for a substantial number of parents to favor the name regardless of the gender of their child.

Why are parents less enthused about giving their sons androgynous names? As is the case for clothing and hair styles—where it is usually females who adopt tastes initially associated with males rather than vice versa—we suggest that there are issues of *contamination* such that the advantaged have a greater incentive to avoid having their status confused with the disadvantaged. In this sense, first names are similar to the caste markers of India in the past—it is far more important for the higher strata to demonstrate their position than it is for those at the bottom of the hierarchy. There is more to be lost for the advantaged and more to be gained by the disadvantaged when customary markers disappear. Likewise, parents may have less difficulty giving their daughter a name that is not sex specific, but find such a choice unappealing for their sons. This is precisely an issue of contamination. The advantaged group has more to lose in symbolic terms when their distinctive features are merged with the less advantaged population. The aversion toward gender neutrality for males, revealed by these naming practices, is substantial and can be understood in terms of the imagery associated with such names.

For similar reasons, asymmetry occurs in other gender areas as well.

Although the terms "sissy" and "tomboy" represent, respectively, deviations from "conventional behavior" such that boys behave like girls and vice versa, Thorne (1993, pp. 115–25) finds that "sissy" is a far stronger form of disapproval. Among college students, she observes, "Unlike the many women who reminisce quite positively about having been called a 'tomboy,' very few men affirm 'sissy' as the core of an earlier (or present) identity" (Thorne 1993, p. 117). Similarly, by fifth grade Thorne reports more stigma attached to boys wanting to participate in conventional girls' activities, such as rope jumping, than there is for girls who move in the opposite direction.

This behavior pattern offers another clue to issues surrounding male withdrawal from occupations that women begin to enter. Obviously, potential monetary and other material rewards are always important factors, and they are not simply symbolic matters. However, this study indicates how imagery may be influencing the choice of androgynous names and, further, how a collective behavior perspective applied to residential segregation may also be relevant in other areas of social life. It is plausible that a perceived threat of symbolic contamination is *one* important factor in the retreat of men from work that begins to be associated with women. This is different, in a very significant way, from saying that the jobs in which women replace men are less prestigious—when "prestigious" refers to income, opportunities, skills, and the like. What we find here is evidence suggesting that gender contamination is a powerful influence even in an area where there are no "bread-and-butter" considerations. It may well be a force influencing the maintenance of gender boundaries in the work world and elsewhere.

REFERENCES

- Alford, Richard D. 1988. *Naming and Identity: A Cross-Cultural Study of Personal Naming Practices*. New Haven, Conn.: HRAF Press.
- Barry, Herbert, III, and Aylene S. Harper. 1982. "Evolution of Unisex Names." *American Name Society* 30 (1): 15–22.
- . 1993. "Feminization of Unisex Names from 1960 to 1990." *American Name Society* 41 (4): 228–38.
- Bell, Wendell. 1954. "A Probability Model of the Measurement of Ecological Segregation." *Social Forces* 32:357–64.
- Bobo, Lawrence, and Camille Zubrinsky. 1996. "Attitudes on Residential Integration: Perceived Status Differences, Mere In-Group Preference, or Racial Prejudice?" *Social Forces* 74 (3): 883–909.
- Buchanan, Barbara, and James L. Bruning. 1971. "Connotative Meanings of First Names and Nicknames on Three Dimensions." *Journal of Social Psychology* 85: 143–44.
- Cohn, Samuel. 1985. *The Process of Occupational Sex-Typing*. Philadelphia: Temple University Press.
- Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, Mass.: Harvard University Press.

- Dunkling, Leslie, and William Gosling. 1983. *The Facts on File Dictionary of First Names*. London: J. M. Dent & Sons.
- Farley, Reynolds, Howard Schuman, Suzanne Bianchi, Diane Colasanto, and Shirley Hatchett. 1978. "Chocolate City, Vanilla Suburbs: Will the Trend toward Racially Separate Communities Continue?" *Social Science Research* 7:319-44.
- Farley, Reynolds, Charlotte Steeh, Tara Jackson, Maria Krysan, and Keith Reeves. 1993. "Continued Racial Residential Segregation in Detroit: 'Chocolate City, Vanilla Suburbs' Revisited." *Journal of Housing Research* 4 (1): 1-38.
- Hanks, Patrick, and Flavia Hodges. 1990. *A Dictionary of First Names*. Oxford: Oxford University Press.
- . 1997. *A Concise Dictionary of First Names*, rev. ed. Oxford: Oxford University Press.
- Jacobs, Jerry A. 1986. "Trends in Contact between Men and Women at Work, 1971-1981." *Sociology and Social Research* 70 (3): 202-6.
- . 1989. *Revolving Doors: Sex Segregation and Women's Careers*. Stanford, Calif.: Stanford University Press.
- Liebertson, Stanley. 1963. *Ethnic Patterns in American Cities*. New York: Free Press of Glencoe.
- . 1980. *A Piece of the Pie: Blacks and White Immigrants since 1880*. Berkeley and Los Angeles: University of California Press.
- Liebertson, Stanley, and Eleanor O. Bell. 1992. "Children's First Names: An Empirical Study of Social Taste." *American Journal of Sociology* 98 (3): 511-54.
- Liebertson, Stanley, and Kelly S. Mikelson. 1995. "Distinctive African American Names: An Experimental, Historical, and Linguistic Analysis of Innovation." *American Sociological Review* 60:928-46.
- Massey, Douglas S., and Nancy A. Denton. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, Mass.: Harvard University Press.
- Peary, Danny, ed. 1978. *Close-Ups: Intimate Profiles of Movie Stars by Their Costars, Directors, Screenwriters, and Friends*. New York: Simon & Schuster.
- Reskin, Barbara F., and Irene Padavic. 1994. *Women and Men at Work*. Thousand Oaks, Calif.: Pine Forge Press.
- Reskin, Barbara F., and Patricia A. Roos. 1990. *Job Queues, Gender Queues*. Philadelphia: Temple University Press.
- Rickel, Annette U., and Lynn R. Anderson. 1981. "Name Ambiguity and Androgyny." *Sex Roles* 7 (10): 1057-66.
- Rossi, Alice S. 1965. "Naming Children in Middle-Class Families." *American Sociological Review* 30:499-513.
- Schelling, Thomas C. 1978. *Micromotives and Macrobehavior*. New York: W. W. Norton.
- Stewart, George R. 1979. *American Given Names*. New York: Oxford University Press.
- Thorne, Barrie. 1993. *Gender Play: Girls and Boys in School*. New Brunswick, N.J.: Rutgers University Press.
- Weekley, Ernest. 1939. *Jack and Jill: A Study in Our Christian Names*. London: John Murray.
- Williams, Christine L. 1989. *Gender Differences at Work*. Berkeley and Los Angeles: University of California Press.
- Withycombe, E. G. 1977. *The Oxford Dictionary of English Christian Names*, 3d ed. Oxford: Oxford University Press.
- Yonge, Charlotte M. 1884. *History of Christian Names*. London: Macmillan.
- Zubrinisky, Camille L., and Lawrence Bobo. 1996. "Prismatic Metropolis: Race and Residential Segregation in the City of the Angels." *Social Science Research* 25 (4): 335-74.

Social Capital at Work: Networks and Employment at a Phone Center¹

Roberto M. Fernandez, Emilio J. Castilla, and Paul Moore
Stanford University

This article argues that a common organizational practice—the hiring of new workers via employee referrals—provides key insights into the notion of social capital. Employers who use such hiring methods are quintessential “social capitalists,” viewing workers’ social connections as resources in which they can invest in order to gain economic returns in the form of better hiring outcomes. Identified are three ways through which such returns might be realized: the “richer pool,” the “better match,” and the “social enrichment” mechanisms. Using unique company data on the dollar costs of screening, hiring, and training, this article finds that the firm’s investment in the social capital of its employees yields significant economic returns.

There is by now a rich empirical literature on the role of social networks in labor markets (for a recent review, see Granovetter [1995], afterword). Almost all of these studies have focused exclusively on the supply side of the labor market, adopting the point of view of job seeker and their social contacts. Numerous studies compare the labor market outcomes of job seekers that obtained their jobs via personal contacts with job seekers that found their jobs by other means (e.g., Bridges and Villemez 1986; Granovetter 1995; Lin, Ensel, and Vaughn 1981). While the supply-side focus has yielded important theoretical and substantive insights, this line of research is incomplete because it largely ignores the demand side of the labor market. Whereas theoretical accounts of the job-person matching process have emphasized the importance of the demand side of the labor market (e.g., Granovetter 1981; Sørensen and Kalleberg 1981), empirical studies of the role of social networks from the employer’s side of the labor

¹ We are grateful for financial support from the Citibank Behavioral Research Council and the Graduate School of Business at Stanford. We have benefited greatly from the comments and suggestions of Jim Baron, Maura Belliveau, Mark Granovetter, Herminia Ibarra, Rakesh Khurana, Maureen McNichols, Trond Petersen, Paul Resnick, Marc-David Seidel, Tony Tam, Ezra Zuckerman, and the *AJS* reviewers. Direct correspondence to Roberto Fernandez, Graduate School of Business, Stanford University, Stanford, California 94305. E-mail: fernandez_roberto@GSB.stanford.edu

market have been rare (for exceptions, see Fernandez and Weinberg 1997; Petersen, Saporta, and Seidel 1998).

In this article, we take an important step toward correcting this imbalance. We focus on the organizational processes at work on the employer's side of the job-person matching process. Specifically, we argue that social networks are deeply implicated in an organizational routine commonly used on the employer's side of the labor market, the practice of hiring new workers via employee referrals. Moreover, we contend that this practice distills elements of a key concept in economic sociology, the notion of social capital (Burt 1992; Coleman 1988).² By its very nature, hiring via referrals is a process that flows through employees' social networks. In seeking to leverage their workers' social networks for their own purposes, employers who use such hiring methods are quintessential "social capitalists." Furthermore, the fact that employers often pay monetary bonuses to their employees for successful referrals suggests that employers view workers' social connections as resources in which they can *invest*, and which might yield economic returns in the form of better hiring outcomes. Understanding the mechanisms at work in this common practice has important theoretical implications for economic sociology, potentially enriching theoretical debates on the embedded nature of economic processes (Granovetter 1985) and the means by which social capital garners returns (e.g., Lin et al. 1981).

Despite the theoretical importance of this common feature of organizational life, the precise social mechanisms at work in hiring via referrals remain obscure. In our current research, we seek to shed light on the referral-hiring phenomenon. We study a research setting that is unusually well suited for identifying and empirically isolating the social processes operating in hiring employee referrals. Within this setting, we analyze unique data on the pool of applicants for entry-level positions at a telephone customer service center of a large bank. We complement these fine-grained analyses by looking at a number of outcomes in several other settings within the bank.

We identify three competing explanations of the referral hiring process. They are the "richer pool" of applicants interpretation, the "better match" argument common within economics, and the "social enrichment" mechanisms emphasized by sociologists. We develop a set of falsifiable hypothe-

² The term "social capital" has been used in a number of ways. For example, Putnam (1993) uses the term to describe features of social organization, such as norms or trust, which facilitate coordinated action. Others (e.g., Burt 1992; Lin et al. 1981) use the term to reflect the instrumental value of social relationships (see Burt [1998] for a review). In this article, we build exclusively on the latter notion of social capital: we assess the instrumental advantages to the firm in using its employees' social networks.

ses that sharply delineate among these accounts. In order to address the richer pool argument, we use data on the pool of applicants to entry-level jobs and directly test whether referrals show evidence of being more appropriate for the job at the application stage. To test the better match interpretation of referral hiring, we study previously unexamined processes involved in screening applicants for entry-level jobs. Ours is the first study to identify theoretically crucial prehire links in the causal logic of the better match argument and to search for empirical evidence of these processes. Finally, we construct a critical test of the social enrichment explanation of recruitment via referrals by examining data on interdependence between referrals and referrers on posthire attachment to the firm.

We have a second important goal for this research. We think the time is right for scholars working in the area of economic sociology to be much more precise in our application of the term "capital" (see Baron and Hannan's [1994, pp. 1122–24] section, "A Plethora of Capitals"). We agree with Burt (1998) that the concept of social capital has become a metaphor for many disparate network processes. If the term "social capital" is to mean anything more than "networks have value," then we will need to demonstrate key features of the analogy to "real" capital. If "social" capital is like "real" capital, we should be able to concretely identify the value of the investment, the rates of return, and the means by which returns are realized. In this article, we offer the literature on social capital an unprecedented level of specificity with respect to the "capitalness" of an important social process by anchoring our analyses of the employee referral practice in *real dollar terms*. Using unique company data on the dollar costs of screening, hiring, and training, we identify the dollar investments that the firm has made by using referrals in their hiring process, we calculate the rate of return on this investment, and we partition the dollar returns across the various social mechanisms by which employee referrals may yield economic benefits to the firm.

WHY HIRE REFERRALS?

Recruiting new employees by means of referrals from current employees is a common practice. The National Organizations Study, a nationally representative sample of employers, shows that 36.7% of employers frequently use employee referrals as a recruitment method (Kalleberg et al. 1996, p. 138; also see Marsden 1994; Marsden and Campbell 1990). Nor is the practice new. Referral hiring was identified as common in the earliest systematic studies of labor markets (Myers and Shultz 1951; Rees and Shultz 1970). More recently, the practice has come to be accepted as a legitimate recruitment tool in the modern human resources/personnel management literature (see e.g., LoPresto 1986). In spite of the widespread

enactment of referral hiring, there is little agreement on how the process works or the ways it produces supposed benefits for employers.

At its most general level, referral hiring is a practice that leverages employees' social ties to provide benefits to the hiring firm. While the various theoretical treatments of the role of networks in hiring often blend these mechanisms (see below), there are five analytically distinct ways by which employers can realize benefits from their workers' networks to their own ends.³

Mechanism 1 is the expansion of the pool of applicants by referrals (Breagh and Mann 1984; Schwab 1982; Fernandez and Weinberg 1997). According to this argument, relying on referrals has the consequence of expanding the firm's recruiting horizon and tapping into pools of applicants who would not otherwise apply (see Rees's [1966] discussion of "extensive" information). Mechanism 2 is based on the tendency for people to refer others like themselves. Since referral ties are likely to be homophilous (Ullman 1966; Myers and Shultz 1951; Rees and Shultz 1970; see also Granovetter 1995),⁴ and since referring employees have survived a prior screening process, such homophily would lead the applicants referred by employees to be better qualified than nonreferred applicants. Indeed, the idea that there is "inbreeding bias" in the referring process is a key component of Montgomery's (1991) theoretical model of social networks in the labor market. Mechanism 3 is reputation protection. To the extent that employees think their reputations within the company will be affected by the qualities of the person they refer, they should only refer qualified applicants (Rees 1966; Rees and Shultz 1970; Saloner 1985; Ullman 1966).

All three of these mechanisms suggest that the pool of referred applicants should be more qualified and more readily hireable than nonreferral applicants (see e.g., Kirnan, Farley, and Geisunger 1989). To the extent that referred applicants constitute a richer hiring pool than nonreferrals, this suggests an important means by which employers can realize returns from using the social capital of their employees during recruitment. Irrespective of which of the three mechanisms are at work, if referrers deliver more appropriate applicants in terms of easily measurable characteristics (e.g., education or work experience), then it would take fewer screens to

³ We are grateful to an anonymous reviewer for helping us to identify these mechanisms.

⁴ This feature has led many to argue that referral hiring is inherently exclusionary (e.g., see LoPresto 1986; Petersen et al. 1998). In other work (Neckerman and Fernandez 1998), we have shown that the effect of referrer-referral homophily is to reproduce the composition of the workforce. In that case, the firm had a racially diverse workforce, and the referral program produced many racial minorities in the referral pool. While referral hiring cannot be counted on to introduce balance into a skewed workforce, it will maintain such balance once it has been achieved.

hire appropriate people from among a pool of referral applicants than it would nonreferral applicants. Since time and resources spent in recruiting are valuable, hiring from a richer pool of applicants reduces the employers' screening costs. We summarize this motivation for referral hiring as the richer pool argument (Fernandez and Weinberg 1997). Mechanisms 1–3 lead to our first hypothesis:

HYPOTHESIS 1.—*Referrals will present more appropriate applications than nonreferrals.*

While mechanisms 1–3 are analytically distinct, all three mechanisms could be operating simultaneously in our data, producing support for hypothesis 1. The reputation-on-the-line process does not rely on homophily: even underperforming employees should worry about risking their reputation with the firm when referring others (e.g., Saloner's [1985] model makes no reference to homophily as a process). Nor does the homophily argument require that referrers target their appeals on particular friends and acquaintances who will not reflect badly on the referrer, as is suggested by the reputation-on-the-line argument. While we do not have measures of the broadening of the recruitment horizon (mechanism 1) and reputation protection (mechanism 3) processes in this study, we can directly assess whether there is evidence of referrer-referral homophily (mechanism 2). The homophily argument leads to our second hypothesis:

HYPOTHESIS 2.—*Among initial applicants, human capital characteristics of referrers and referrals will display above-chance levels of homophily.*⁵

Mechanism 4 by which referral hiring might help employers focuses on the informational advantages of hiring via referrals (Rees and Shultz 1970; Ullman 1966). The information advantage theory of referral hiring argues that ties between people also provide a conduit for information to flow between the employer and the applicant. In addition to locating particularly appropriate potential candidates who otherwise might not apply (i.e., mechanism 1, Rees's [1966] "extensive" margin), the referrer adds information along the "intensive" margin (Rees 1966). In this model, the referrer passes on extra, hard-to-measure information about the worker's qualities to the employer (e.g., personality attributes or attitude). On the applicant's side, the referrer also deepens the applicant's information about the job (e.g., the informal rules governing performance) such that

⁵ While many past studies have demonstrated a relationship between the characteristics of job contacts and people hired via networks (e.g., Lin et al. 1981; De Graaf and Flap 1988; Neckerman and Fernandez 1998), these studies have all looked at the degree of similarity between hires and their contacts. Such studies cannot control for possible selection biases involved in screening (see Fernandez and Weinberg 1997). To our knowledge, ours is the first study that addresses the question of prehire homophily, i.e., between applicants and their referrers.

referrals have a better sense than nonreferrals of "what they are getting into."⁶ Since this information may not be available by any other means, such inside information has the effect of allowing both employers and applicants to make better, more informed decisions than nonreferrals. Advanced knowledge of what the job is like is alleged to reduce frictions between the worker and the job along these hard-to-measure dimensions, thereby improving the match between the person and the job. Studies in this tradition have suggested that referral hires should have higher starting wages (Simon and Warner 1992), slower wage growth (Simon and Warner 1992), and lower turnover (Corcoran, Datcher, and Duncan 1980; Datcher 1983; Simon and Warner 1992; Sicilian 1995) than nonreferrals. While there are a number of nuances regarding exactly how the extra information associated with referrals is harvested by the various actors involved in the hiring process (see hypotheses below), this information-improves-match mechanism is common to economic treatments of hiring via referrals (e.g., Rees and Shultz 1970; Simon and Warner 1992). We refer to this as the better match account of referral hiring.

Despite the fact that the better match argument's key causal mechanisms exclusively operate during the screening process, there has been no empirical work directly examining the information-related mechanisms posited to affect screening in the matching story. While there have been numerous theoretical models exploring the implications of assuming that various parts of the information exchange process work as described above (e.g., Montgomery 1991; Saloner 1985), empirical research on referral hiring in the matching tradition has "black boxed" the hiring process and has focused solely on the *posthire* implications of referrals' better matches. Instead of *inferring* the existence of extra information from post-hire outcomes, we examine data on the hiring process for *direct* evidence of extra information associated with referrals.⁷ We have assembled unique, fine-grained data on the screening and hiring process (see below) in order to empirically test for evidence of the information-related pro-

⁶ Wanous (1980) is most explicit about this particular mechanism. He argues that the information that referrals pass on about the more tacit features of the job serve much the same functions as a "realistic job preview." Similarly, Jovanovic (1979; see also Mortensen 1988) has argued that jobs are "experience goods" in that it is hard to judge how much one would like a job before taking it.

⁷ Although it constitutes less direct evidence of matching than the extra information mechanisms we focus on here, we will also report results for the traditional posthire indicators of referrals' better matches, i.e., higher starting wages (Simon and Warner 1992), slower wage growth (Simon and Warner 1992), lower turnover (Corcoran et al. 1980; Datcher 1983; Simon and Warner 1992; Sicilian 1995), and the time path of turnover. As we discuss below, these analyses will be useful for distinguishing among various versions of the matching argument.

cesses that are posited by the matching story to operate during screening. Ours is the first study to open this black box and directly test the prehire information predictions of the matching model. We examine whether referrals show evidence of having knowledge of key information about the job at various stages of the screening process. Such information has traditionally been treated as an unobservable part of the screening process. Moreover, we study the various ways that employers may get extra information about referral candidates. Since it is employers who design and execute such recruitment and screening processes, this analysis would provide key insight regarding exactly *how* employers leverage their employees' social capital during screening.

Beginning with the applicants' side of the process, if referrers deepen the candidate's information about the job, then potential applicants who are referred should be better able than nonreferrals to decide whether the job will be of interest and worth pursuing (see Rees and Shultz 1970; Granovetter 1995; Wanous 1980).⁸ Consequently, among those who choose to apply, referral applicants should be particularly well informed about the nature of the job when compared with nonreferrals. This motivates our hypothesis 3:

HYPOTHESIS 3.—*At application, referrals will have more information about the nature of the job than nonreferrals.*

Especially if referrals have decided that the job is worth pursuing, referrers might provide referrals with inside information on how the hiring and screening process works at the company. Such information would lead applicants to change their application behavior so as to capitalize on this information. In particular, referrals might attempt to time their application for periods when the chances of being hired are more favorable (see Fernandez and Weinberg [1997] for another test of this hypothesis using different data from another unit of the same bank). This leads to hypothesis 4:

HYPOTHESIS 4.—*Referrals will better time their applications than nonreferrals.*

⁸ This introduces the possibility that selection bias (Winship and Mare 1992) will affect our test of the better match theory. All of the people who applied to this firm are self-selected in the sense that they had enough interest in the firm to pursue a job there. However, to the extent that referrals have jobs explained to them, but opt out and decide not to apply for the job, then the survivors among the referrals may be more selected than the nonreferrals. Differential selectivity is predicated on the idea that, compared with nonreferrals, referrals have extra knowledge of the job given to them by their referrers. By focusing on survivors who actually apply in this case, differential selection bias would work in the direction of overestimating the extent to which referrals have extra knowledge of "what they are getting into." In this case, our conclusions are not threatened by this differential selection (see below).

Another consequence of referrers passing on extra information about the job to referrals is that referrals should be presold on the job (Ullman 1966). If the better match theory is correct, then referrals should show less equivocation about accepting job offers than nonreferrals. This leads to hypothesis 5:

HYPOTHESIS 5.—*Referrals will accept job offers at a higher rate than nonreferrals.*

Finally, if the better match account of referral hiring is valid, referrals should report having superior knowledge of the job content and tasks than nonreferrals at hire. This leads to hypothesis 6:

HYPOTHESIS 6.—*Referrals will report greater levels of understanding of job content than nonreferrals.*

In hypotheses 3–6, the emphasis is on testing for evidence of referrals having superior information about the job than nonreferrals. However, the matching story contends that similar processes should work on the employers' side as well. In order to get benefits from better match processes, employers, too, should show evidence of having and using better information about referrals than nonreferrals during screening.

There are two principal ways in which employers may harvest extra information about referral candidates. First, employers can rely on relatively indirect means of information gathering. Employers could utilize the "upstream" information that could be available via the tendency of people to refer others like themselves (the homophily principle; see hypothesis 2) during their screening decisions. Montgomery's (1991) theoretical model argues that employers are aware of homophily in referral networks and, consequently, use the characteristics of the referrer as an "upstream" signal of the qualities of the referred applicant (see also Miller and Rosenbaum 1997; Ullman 1966). This argument would predict that the *referrers'* characteristics should affect the firm's screening decisions, independent of the applicant's characteristics. This yields hypothesis 7:

HYPOTHESIS 7.—*Conditional on hypothesis 2 being supported, and controlling for applicant's human capital characteristics, recruiters should use the characteristics of referrers when screening referral applicants.*

However, employers can also turn to very direct methods of information gathering. More specifically, employers can inquire directly with the referrer about characteristics of the referral candidate. This yields hypothesis 8:

HYPOTHESIS 8.—*Recruiters should contact referrers when screening referral applicants.*

The key difference between the richer pool and better match theories lies in assumptions about the behavior of the referrer. This difference has very important implications for understanding how it is that employers can realize returns on social capital. If the richer pool process were the

only one to be operating, then referrers might produce applicants who are no better informed than nonreferrals about the more tacit features of the job but are more appropriate than nonreferrals in terms of easy-to-screen-for, formal qualifications for the job. If referrals' advantages were on easily screened, formal characteristics, then once recruiters apply their screen, we should see no differences between referrals and nonreferrals in their posthire behavior (e.g., turnover). This suggests that firms may save on screening costs, since it would take fewer screens on formal qualifications to fill jobs from the referral pool than the nonreferral pool; but referrals and nonreferrals would not show any posthire performance differences.

If the better match process operates, but not the richer pool process, then referrals might be better informed than nonreferrals about tacit features of the job but no better than nonreferrals in terms of formal qualifications. In this circumstance, there would not be any cost savings in screening, since it would take just as many screens to filter out formally unqualified referral applicants as nonreferral applicants. But, there would be savings on posthire consequences, namely lower turnover. Once candidates have been screened on formal qualifications, then among the survivors, referrals should be better informed than nonreferrals about the more informal characteristics of the job. Since referrals would have a better sense of what the job entailed than nonreferrals, fewer referrals than nonreferrals would conclude upon experiencing the job that the job is not for them and leave. The better match theory, then, posits a second source of returns to employers for their social capital investment: savings due to referrals' lower turnover (e.g., Halcrow 1988; Kirnan et al. 1989; Saks 1994; Wanous 1980).⁹

The last mechanism (mechanism 5) by which referral hiring yields ad-

⁹ Some versions of the matching model have incorporated richer pool arguments as part of their matching explanations. For example, the homophily and reputation protection arguments have been used to infer that referrals are better matched to their jobs, under the presumption that referrer's search produces not only applicants who are higher quality on formal criteria, but also produces better matched applicants on the more tacit, hard-to-screen-for characteristics (Simon and Warner 1992). To the extent that referrals are higher quality on the hard-to-screen-for characteristics, then referral status acts as a signal: referral status is associated with superior posthire outcomes but is hard to mimic by those lacking those superior, hard-to-screen-for characteristics (Spence 1974). It is plausible that these processes often work together, since we would expect referrers to pay attention to both formal qualifications and more tacit skills when referring a worker. As we show below, the "referral status as signal" version of the matching story yields some predictions that are quite different from the richer pool argument. Although we are hampered by the fact that the better match models that incorporate the richer pool processes work in the realm of unobservable factors, we will attempt to distinguish between the better match models that rely on superior information (i.e., mechanism 4) from versions of the matching argument that fold in the richer pool processes (mechanisms 1-3).

vantages emphasizes social processes that occur post hire. In this model, the connection between the new hire and the job is enriched by the existence of a prior friend or acquaintance that might ease the transition to a new job setting. As such, referrers serve as a kind of naturally occurring mentoring system, aiding newcomers in the organizational socialization process (Reichers 1987; Sutton and Louis 1987). We call this the social enrichment argument for referral hiring.

While the social enrichment process would work independently of the better match process, it can produce similar posthire behavior by the referred hire. Both processes may lead to outcomes that are desirable from the firm's perspective, for example, lower turnover (Wanous 1980), lower absenteeism (Taylor and Schmidt 1983), and higher productivity (Swaroff, Barclay, and Bass 1985). However, there are important distinctions between these two accounts in the ways in which these benefits are realized. In the better match story, applicants are treated as socially isolated actors. Referrals yield benefits by improving the firm's ability to pluck the right individuals for the job from the pool of applicants. Likewise, the extra information about the job provided to referrals by referrers helps applicants to make better-informed decisions about whether to apply for the job than nonreferrals.

In contrast, the social enrichment account has referrers actively changing the relationship between the firm and the new hire. It is the presence of social ties that directly benefits the firm. Social relationships create new human capital in the referred person, post hire, via referrers' assistance with tacit aspects of the work environment and informal training on job tasks. The assistance that referrers offer can also improve new hires' firm-specific human capital since such assistance is occurring within the context of particular firms. Moreover, from the referred applicant's point of view, the relationship with the referrer may have value in itself, which improves the new hire's workplace experience. This can increase satisfaction, commitment, and attachment to the firm, saving the firm the expense of the costs of training of replacements due to turnover.

While the social enrichment process can work in a manner that is beneficial from the employer's perspective, it is also possible for this process to work *against* the interests of the employer. In contrast with the matching account where more information is unambiguously good from the employer's point of view, the social enrichment story suggests that employers might experience both costs and benefits when relying on referral hiring. For example, Blau (1985) showed that nurses who were socially integrated had poorer attendance records than socially isolated nurses. He interpreted this as reflecting a tendency for nurse's friends to cover for one another. Also, in Bailey and Waldinger's (1991) account of hiring in garment trades, the fact that workers would not help new hires with on-the-

job training *unless* the new hire was someone already known to the existing workers has the effect of severely limiting employer's hiring discretion. (See Grieco [1987] for another account, where employees' social networks do not necessarily work in the employer's interests). The fact that employers' returns via the social enrichment process are uncertain suggests that social capital in this form shares another feature with financial capital. As an investment with a downside potential, social capital involves *risk*.

The social enrichment account for referral hiring is mute on the subject of referrals' screening advantages. Evidence of either referrals being better at application (as predicted by the richer pool theory) or of extra information during screening (consistent with the better match theory) would not contradict the social enrichment account. The social enrichment account, however, emphasizes another reason why employers may hire via referrals: posthire relations between referrals and referrers affect new-hires' attachment to the company. While the better match theory sees no role for posthire social relations in affecting referrals' behavior, the social enrichment model posits that there will be interdependence between the posthire attachment of the referrer and referral. This leads to our final hypothesis:

HYPOTHESIS 9.—*Referrals' turnover should be affected by the turnover of the referrer.*

If there is interdependence between posthire attachment to the firm for referrals and the people who referred them, then employers can anticipate and predict such posthire social processes. Such interdependence would constitute yet another mechanism by which employers might reap benefits from their workers' social capital. But as we noted above, the fact that such social processes are not totally under the employer's control implies that employers might experience costs as well as benefits from referral hiring. Like all forms of capital, social capital can be squandered or invested wisely.

RESEARCH SETTING

We studied the hiring process for an entry-level position at a large, mid-western phone center, within a large, globally diversified financial service institution. The job we study is the Phone Customer Service Representative (PCSR). This is a full-time, hourly position which duties consist of answering customers' telephone inquiries about their credit card accounts. New hires into this position are given two months of classroom and on-the-job training before working on the phone. PCSRs are trained in balancing the etiquette of customer service interactions with accuracy, speed, and efficiency while processing phone calls. PCSRs can expect to handle

up to 5,000 phone calls per month. Phone calls are often monitored by managers to insure that PCSRs' courtesy and accuracy goals are being met (see Attewell [1987] for a discussion of these surveillance technologies).

We studied records of the phone center's hiring activities during two years (from January 1995 through December 1996). The phone center's human resources department (hereafter PCHR) tracked more than 4,100 external employment inquiries for PCSR jobs over this two-year period. In order to address the posthire consequences of the recruitment practices, we also examined the turnover and wage histories of individuals hired during the two-year hiring window until January 1998.

The main focus of this study is this one site, but, where possible, we supplement these analyses with data from three of the bank's other phone centers located in different cities across the United States. These other sites do not keep the detailed data necessary to examine hypotheses regarding the prehire phase of recruitment. However, we utilize data on the posthire consequences of referral hiring (i.e., turnover and wage growth) at these other sites to replicate the analyses in the midwestern site. We can also replicate some portions of both the prehire and posthire analyses with similar data on entry-level positions collected from a unit of the retail bank (see Fernandez and Weinberg 1997). We report the results for these other sites at appropriate points in the article.

Because a major focus of this study is to unpack economists' better match theories of referral hiring, we have sought to mount a conservative test of such theories. Consequently, we have designed the study to approximate as closely as possible the economists' neoclassical labor market. First, we chose as our main study site a very competitive labor market. During the entire period of our study, the local unemployment rate remained below 4% (labor market conditions in other sites were somewhat less competitive but were also experiencing rapidly declining unemployment). Under such conditions, one would expect firms to seek economic advantages in the labor market in as many ways as possible, including exploiting any screening advantages associated with the use of referrals. Second, we studied entry-level jobs in order to observe the firm's screening behavior under conditions of maximum exposure to the external market. The strategy of studying lower-level employees has the added advantage of controlling on selection what would otherwise be a major competing argument for referral hiring. Some past studies (Pfeffer 1989; Ferris and Judge 1991) have argued that people often hire others who will support their agendas in political fights within the organization. As entry-level employees, these workers are very unlikely to be involved in organizational politics involving hiring. Other scholars (e.g., Marsden and Campbell 1990) have argued that personal contacts might matter more in hiring for higher-status rather than lower-status jobs. From this per-

spective as well, our study is a conservative test of the role of personal contacts in the hiring process.

The final theoretical advantage for looking at entry-level jobs in this setting is perhaps the most important. To work at a phone center, fielding up to 5,000 phone calls a month in a high-pressure, highly structured environment where work is closely scrutinized, demands a set of skills for which it is relatively difficult to screen. As one of the PCHR recruiters put it, "If you look at the paper requirements to do this job, they are not heavy. You don't need a degree to do this job. This job can be done by people with average intelligence. People rarely leave because they can't do the job. The real question is are you going to be able to deal with the work environment." More than most jobs, this job needs to be experienced in order to determine how well one will like it. According to the better match account, the information about the job that referrers provide referrals is especially valuable when more tacit aspects of the job are especially salient. By this reasoning, the environmental features of the PCSR job make this a setting in which we are very likely to observe better match processes at work.

In addition to these theoretical considerations, the phone center offered a number of practical advantages for this research. PCHR keeps virtually complete databases on recruitment for phone service representative jobs, which has allowed us to track applicants' movement through every phase of the hiring process. In addition to these computer databases, PCHR keeps paper files on each applicant, including a standardized application form and all material offered by the applicant in support of the applications (e.g., a resume). These paper files allow us to code crucial data on applicants' education, work history, and other human capital characteristics, as well as data on applicants' prior knowledge of the job.

This setting has several other features that make it ideal for addressing research on social networks and hiring. First, because PCHR traces employment inquiries from initial contact, through interview, to actual hire, we are able to treat hiring as a multistage process in which referral ties may play different roles at the various stages. Unlike studies that start with a set of hires, this design feature allows us to address the question of selection bias when studying the impact of referral ties in the hiring process (for a discussion of this issue, see Fernandez and Weinberg 1997). Hence, compared to existing research, we are in a much better position to identify the precise influence of referral ties at each step in the selection process for this job. Furthermore, the PCHR represents an exceptional location at which to study the effect of referrals on posthire processes as well. These data allow us to examine the duration of the employment and salary history of individuals who were hired during 1995 and 1996.

Second, these data contain information on key variables of interest at

application: the presence or absence of a referral tie, as well as the identity of the *referrer*. Unlike past research where data on the characteristics of the job contact are observed only among hires (e.g., Corcoran et al. 1980; Lin et al. 1981; Neckerman and Fernandez 1998), we are able to link referrers to job *applicants*. Consequently, we are in the unique position of being able to study the influence of referrers' characteristics on referrals' chances of being hired.

A third major advantage of this research is the fact that we have been able to learn PCHR's screening criteria (see the section on procedures below). Consequently, we can more precisely specify the set of appropriate individual control variables that affect labor market matching. We can therefore consider the extent to which an applicant's referral status is a proxy for other characteristics that might make the applicant desirable to the PCHR recruiters. Sparse controls for human capital and other characteristics raise the risks of biased assessments of the role of social contacts in the hiring process (Fernandez and Weinberg 1997).

Although we are studying only one firm, it is worth noting that our firm's hiring practices are not particularly distinctive. Specifically, this firm is unremarkable with respect to the screening and interview procedures it uses (see Kalleberg et al. 1996, p. 138). Furthermore, like our firm, firms that pay current employees for employee referrals appear to be following an accepted modern personnel practice (see, e.g., Halcrow 1988; LoPresto 1986; Stoops 1981, 1983). The question of generalizability of the results remains; however, we have taken some steps toward addressing this issue in this article by analyzing less complete posthire data in three other phone centers and in the retail bank located elsewhere in the United States. While all studies carry the burden of the time and place in which they have been conducted, it is important to realize that the comprehensive information on job matching we have available here is unique and is not available through any other study. Indeed, for many of the hypotheses developed below, no empirical evidence at all has ever been offered. It is impossible to distinguish among the different theoretical mechanisms through which social ties might work in the hiring process without the kind of fine-grained data we analyze here.

PROCEDURES

We combined data from a number of sources in order to address the hypotheses developed below. As a part of their standard operating procedures, PCHR professionals record in a computer database the recruitment source for every employment inquiry at initial contact with the bank, and they keep this information regardless of whether the prospective applicant proceeds to the next phase of screening or not. If the applicant survives

a paper screen by PCHR staff, they are contacted and briefly interviewed either in person or by phone. Applicants who survive this phase of the screening are then sent on for another interview with two hiring managers who have the final say about extending the candidate a job offer.

During 1995–96, PCHR received 4,165 external applications for PCSR positions, and 60.1% of those applicants were sent on for an interview with the hiring manager. Only 8.6% (357) of the applicants were offered a job, and 7.8% (327) of the original applicants were hired. In order to address posthire processes, we tracked 325 of these hires from hire until January 1998, when we ended the data collection.¹⁰ Almost half (162 of 325) of the hires are right censored, that is, they were still with the company at the end of our study. For hires, the days of tenure with the phone center range from a minimum of 3 days up to a maximum 1,104 days, with a median of 480 and a mean of 528 days.

One of the most unique features of these data is the fact that we have been able to link referrers with their referrals at the application phase. More than one-third of the applications (37.1%, or 1,546) were referrals, and slightly less than two-thirds were nonreferrals (62.2%, or 2,594).¹¹ PCHR coded the name of the referrer in their tracking database.¹² 1,223 referrers produced these 1,546 referrals. From company data sources, we identified employment records for 97.5% (1,192) of the referrers. The number of referrals per referrer vary from 1 to 6, although the vast majority of these referrers referred only one (79.7%) or two (15.8%) individuals. We coded three key variables of interest for referrers: wage at the time they referred the applicant, tenure with the firm, and years of education.

We interviewed PCHR staff in order to determine their screening criteria. They informed us that they saw their main role as one of saving the hiring manager's time. They would "send [the hiring manager] people we know they will want to hire, as well as people we are not sure they will not want to hire" (field quotation from a PCHR recruiter).¹³ With respect

¹⁰ We dropped one case because we could not identify its recruitment source; we could not locate posthire records for the other case.

¹¹ We could not identify the recruitment source for 25 applications. We dropped these cases from all subsequent analyses.

¹² There is also a line on the employment application that explicitly asks the applicant to list the name of the referrer. Referring employees are paid \$10 if the people they refer are interviewed and \$250 if the referral is hired and survives a 30-day probation period. This creates an important incentive for referring employees to ensure that applicants list them accurately as their referrer. As we discuss below, this referral bonus also constitutes the firm's social capital investment in the social networks of their employees.

¹³ Developments after the field period of our hiring window lend support to this conception. In an effort to further save recruiting time, PCHR staff have been given final authority to hire on behalf of line managers.

to screening criteria, the PCSR job involves a lot of customer interaction, so PCHR screen people based on verbal and "soft" interpersonal skills, which they glean from short interviews or phone interactions.¹⁴ PCHR recruiters also place relatively high weight on an applicant's job history when screening applications. In light of the customer service aspects of the job, recruiters look particularly for people with prior customer service experience. They are also quite concerned about work attitudes and tend to look for applicants who they think will be reliable employees. This leads them to prefer applicants who are currently employed and who have had some previous work experience. Recruiters are also quite concerned about the costs of turnover, and, therefore, are less impressed by people who have changed jobs a lot during their work histories. In addition to these work history factors, PCHR recruiters look for evidence of basic keyboarding and computer skills on the application.

PCHR recruiters are also concerned about applicants who are "overqualified" for this entry-level position. More specifically, candidates who report higher wages in their previous job than the starting wage at the phone center are looked upon with some skepticism. As one PCHR recruiter put it, "Our past experience has shown us that, unless they have very special circumstances, [people with high wages in their previous jobs] are very likely to get dissatisfied. . . . They overestimate the extent to which they can handle the cut in pay, get dissatisfied, and leave on us." Consequently, applicants who report high wages are deemed *less* appropriate than are those who report wages more in line with those being offered at the phone center. Compared with work experience, PCHR recruiters place less weight on formal education for these entry-level jobs. Although preferred, a high school diploma is not necessary to continue to be given serious consideration. While they do not place much weight on educational criteria in general, PCHR recruiters are concerned that highly educated people might be using these jobs as a platform to look for better employment and, consequently, might leave abruptly. Similar to what we found in the western region of the retail bank (Fernandez and Weinberg 1997), PCHR recruiters say they consider very highly educated applicants as overqualified for these entry-level jobs.

¹⁴ Because of the large number of these contacts, these interviews are not well tracked by PCHR. We know that a hiring manager has interviewed all candidates who receive job offers. We also know that everyone who has been interviewed by a hiring manager has been screened and interviewed by PCHR. For those candidates who are not sent to the hiring manager for an interview, we cannot be sure whether they simply failed a paper screen by PCHR or whether they were rejected after a short screening interview. Consequently, we are forced to treat the interview phase as a joint effort on the part of PCHR and the hiring manager. This is consistent with their own description of their role, in that they seek to mimic the screening decisions that hiring managers would make downstream.

There is one criterion that absolutely disqualifies applicants. The application form asks whether the applicant has ever been convicted of a "breach of trust"; applicants responding "yes" are eliminated from further consideration since regulatory agencies will not allow banks to hire such people into PCSR positions. Breaches of trust include shoplifting, embezzlement, forgery, fraud, and writing checks with insufficient funds. Even if the applicant has answered no to the breach of trust question, PCHR procedures are to repeatedly ask ("some would say badger"; field quote from PCHR recruiter) whether the applicant has been convicted of a breach of trust. All hires are required to undergo expensive fingerprinting and background checks. If they come back showing a conviction, the bank is required to let the new hire go. Especially in light of the very high costs of training new hires (see below), PCHR staff work diligently to screen out such individuals.¹⁵ While there may be other features of an application that occasionally catch a recruiter's eye, these factors are the ones for which PCHR recruiters systematically screen.

We coded the paper files to reflect PCHR's concerns. We located paper files for 97.2% (3,998) of the 4,115 applications for which we could identify a recruitment source and that did not report a breach of trust. Based on our review of the paper files, we recorded a number of variables that measure applicants' human capital. We coded years of education, months of experience in the financial services industry, months of experience outside the banking industry, and, more relevant to the particular position, months of customer service experience. We also coded a dummy variable for whether the person was employed at application. As measures of turnover propensity (recall that stability is of great concern to PCHR recruiters), we coded the number of previous jobs listed on the application, months of tenure with the last firm, and wages of the last job. We looked for evidence of computer experience among the application materials and created a dummy variable for the presence of these skills (the applications had a line specifically asking applicants to supply such information). Finally, as a control, we coded gender from the names listed on the application for use as a control variable. A few records (28, or 0.7%) had ambiguously gendered names; we coded these as missing on gender.

For hires, we coded the dependent variables in the posthire analysis: duration in the company (with exact dates for those who were terminated) and salary changes during their tenure with the firm. We also obtained questionnaire data on a sample of 233 (71.2%) of the hires that occurred

¹⁵ Out of the 4,165 applications for PCSR jobs during our study, we found only 25 applications that had checked "yes" to the breach of trust question, and none of these were interviewed, offered a job, or hired. Because they have no chance of progressing through the hiring process, we have dropped these 25 cases.

during our hiring window.¹⁶ These data were collected during the second week of training and contain important information regarding what applicants know about the PCSR job. The survey asked new PCSR hires to rate their understanding of what the position responsibilities and job content (for PCSR job) was *prior* to accepting the position (see below).

ANALYSES

We address these hypotheses by examining empirical evidence on the hiring process at the midwestern phone center. Wherever possible, we present evidence based on multiple measures when addressing the hypotheses. Where appropriate, we supplement the analyses of the PCHR with data collected from three other phone centers.

Mechanisms 1–3: Referrals Present More Appropriate Applications

Hypothesis 1 posits that mechanisms 1–3 would lead referrals to present more appropriate applications than nonreferrals. Using PCHRs suggestions of what they screen for, we examined nine variables measuring applicant's work history, computer skills, and education from the paper applications to test this prediction.

We begin by comparing the means for referrals and nonreferrals on individual variables (table 1). Univariate *F*-tests for differences in means show that referrals are more likely to be employed at the time of application, show longer tenure with their previous employer,¹⁷ and are more likely to have had fewer jobs (one-tailed tests, $P < .001$, $P < .01$, and $P < .05$, respectively). Referrals also have had more months of customer service experience, although this difference misses being significant at the 5% level ($P < .059$, one-tailed test). Differences between referrals and nonreferrals on months of experience in financial services, months of experience outside financial services, and computer skills are not statistically reliable in the univariate tests.

Because recruiters avoid "overeducated" and "overpaid" applicants, we needed to develop measures of overqualification on these dimensions. In

¹⁶ The survey responses were made anonymously, and, consequently, cannot be matched to other data we have on these individuals.

¹⁷ Small numbers of applicants reported being new labor market entrants who have never had a job before, i.e., 2.4% of nonreferrals and 1.7% of referrals. We coded tenure on the last job and wage on the last job as zero for people who have not had a previous job. Therefore, the tenure and wages differences in the MANOVA (reported below) are conditional on the number of previous jobs being one or greater. None of the substantive results change if we exclude those who have never had a job from the analyses we report here.

TABLE 1

MEAN BACKGROUND CHARACTERISTICS OF REFERRED AND NONREFERRED APPLICANTS

	MEANS		UNIVARIATE <i>F</i> -TEST
	Referrals	Nonreferrals	
Months of:			
Customer service experience	35.073	32.445	2.436
Banking experience	2.014	2.126	.041
Nonbanking experience	66.592	65.394	.262
% overeducated (greater than 16 years)	2.324	3.409	2.862
Computer skills790	.786	.062
Working at time of application630	.469	76.041***
Number of previous jobs	3.230	3.304	3.739*
Tenure on last job (in days)	759.012	675.701	4.484**
% high wages on last job (greater than \$8.25)	24.220	27.070	2.989*
<i>N</i>	1,119	1,936	
Wilks's lambda value ^a971		
Multivariate <i>F</i> -test	10.157***		
<i>df</i>	9 and 3,045		

^a Tenure on the last job and wage on the last job are coded as zero for people who have not had a previous job. The substantive results do not change if we exclude those who have never had a job from the MANOVA.

* $P < .05$.

** $P < .01$.

*** $P < .001$, one-tailed test.

preliminary analyses, we compared the distributions of education and wages on last job for referrals and nonreferrals. When measured at the mean, referrals appear to be slightly less educated than nonreferrals (13.700 vs. 13.893), a difference which is not in the predicted direction. However, this is due to the education distribution for nonreferrals having a long right tail of highly educated people since the medians for the two groups are identical (14 years). This is consistent with referrals having fewer "overqualified" candidates than nonreferrals. To examine this overqualification difference, we compared the rates at which referrals and nonreferrals reported greater than 16 years of education. Only very small percentages of such candidates applied to either pool. While there are fewer overeducated applicants in the referral than nonreferral pool (2.3% vs. 3.4%), the univariate test shows that this difference is not statistically reliable at the 0.05 level ($P < .091$, one-tailed test).¹⁸

¹⁸ In preliminary analyses, we experimented with various nonlinear specifications of the education variable (education-squared, various splines; see n. 28) to look for evidence of overeducation. While referrals are somewhat less overeducated than nonre-

With respect to wages on the previous job, referral applicants earned lower wages on their last job than nonreferrals when the difference is measured at the mean (\$7.18 vs. \$7.47; $P < .01$, one-tailed test). The median wage for both groups, however, is identical: \$6.50. Here, too, nonreferrals have a longer right tail of overqualified applicants than referrals. This is seen clearly in figure 1, which plots the percentile distribution of wages on last job for nonreferrals against the percentile distribution for referrals. The 45-degree line shows a baseline of what the plot would look like if referrals and nonreferrals were to have identical wage distributions. The distributions for referrals and nonreferrals for the observed data closely track the 45-degree line up to \$8.00, which marks the 73rd percentile for both nonreferrals and referrals. However, above that point, the line for the observed data begins to diverge upward from the 45-degree line. This indicates that the nonreferral pool contains a greater proportion of highly paid applicants than does the referral pool. As a conservative definition of overqualification with respect to wages on their last job, we coded the percentages of referrals and nonreferrals that reported wages greater than the non-negotiable starting wage for PCSRs, \$8.25 per hour.¹⁹ Of referrals, 24.2% were overqualified in this way, compared with 27.1% of nonreferrals, a difference that is statistically significant in univariate tests ($P < .042$, one-tailed test).

Hypothesis 1 implies that these variables will form two distinct profiles, one for referrals and one for nonreferrals. While the last column of table 1 showing univariate F -tests is informative, looking at each variable separately is inconsistent with the notion of a profile. In order to test hypothesis 1, we also reported the multivariate test of whether application source (referral vs. nonreferral) is statistically independent of the application profile. We find strong support for hypothesis 1: recruitment source significantly distinguishes between the joint distribution of the nine measures of applicant's background ($P < .0001$; Wilks's $\lambda = .971$; F -test = 10.157; $df = [9, 3,045]$). While we did not have access to background data as rich as we have available here, these findings replicate those of the entry-level jobs we studied in a retail bank (for details, see Fernandez and Weinberg [1997]).

referrals when using these nonlinear education measures, these patterns are never statistically reliable in univariate tests.

¹⁹ From the perspective of hypothesis 1, this definition is conservative because the distributions for referrals and nonreferrals diverge *more* dramatically the further one goes into the tails of the distributions. We experimented with definitions of overqualification, which required a more extreme departure from \$8.25 (i.e., \$9.00–\$15.00); as fig. 1 reveals, setting the overqualification threshold higher increases the extent to which nonreferrals are overqualified compared with referrals.

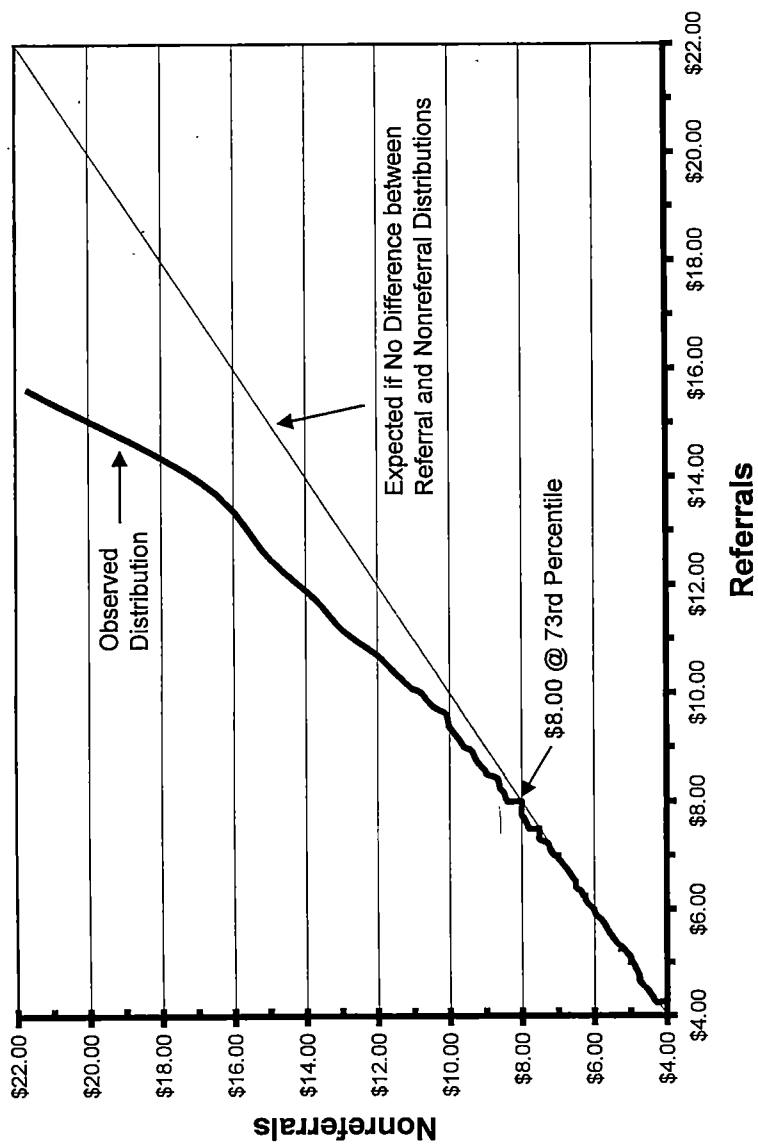


FIG. 1.—Percentile distribution of applicants' wages on their last job by referral status

Mechanism 2: Homophily

We address hypothesis 2 by selecting three characteristics that are likely to reflect referrer's human capital and success with the company, and therefore, are most likely to offer the employer upstream information about the quality of the applicant. We measured referrer's years of education, hourly wage, and days of tenure with the firm, all at the time they made the referral. We also coded the sociodemographic characteristic that is easily discernible from the applications: gender. For referrals, we coded the applicant's gender, years of education, hourly wages, and days of tenure with their current employer. If the applicant was unemployed, we used their wages and tenure from their last job. For the purposes of the homophily tests, we excluded applicants who had never been employed.

As we mentioned above, ours is the first study to address the question of homophily between *applicants* and their referrers. As a standardized measure of homophily across different characteristics, we calculated the Pearson correlation between referrer's characteristics and the characteristics of the referral. We also calculated an unstandardized measure of the magnitude of the variation between referrers and referrals by computing the mean of the absolute value of the difference (hereafter MAVD) between referrers' and referrals' values on each of the four variables.²⁰ Because larger numbers indicate greater differences, the MAVD is a negative measure of homophily.

The first column of table 2 reports the homophily measures. Consistent with the idea that people tend to refer people like themselves, the correlations for all four background variables are positive. However, all four correlations are quite modest, especially the correlation on past wages. While small in magnitude, the large number of dyads (see table 2, col. 2) insures that all of these correlations are statistically reliable when assessed by traditional significance tests. For tenure, education, and gender, the *P*-value is less than .001; for wages, $P < .011$ (one-tailed tests). The size of the correlation on wages is understandable when we consider that there is likely to be limited dispersion among applicants since the PCSR job is an entry-level job and is unlikely to attract many highly paid people (indeed, recruiters expressed skepticism of highly paid applicants for this job). The MAVD measures show the magnitudes of the differences between referrers and referrals in the metric of each of the background variables. For tenure, referrers and their referrals differ by an average of 1,376 days, while their average difference on wages is \$3.96. Referrer-referral dyads differ by 1.77 years of education; 33% of these dyads are cross-gender.

²⁰ We are grateful to an anonymous reviewer for this suggestion.

TABLE 2
TESTS OF HOMOPHILY IN REFERRER-REFERRAL DYADS FOR FOUR CHARACTERISTICS

	EXPECTED VALUES BASED ON RANDOM BASELINE ^a				TESTS VERSUS RANDOM BASELINES ^b			
	Observed Homophily ^c (1)	Number of Dyads (2)	All Applicants (3)	Nonreferrals (4)	Referrals (5)	All Applicants (6)	Nonreferrals (7)	Referrals (8)
Tenure (in days):								
Correlation112	1,108 ^d	0	0	0	$P < .001$	$P < .002$	$P < .002$
MAVD	1,376		1,440	1,446	1,436	$P < .018$	$P < .009$	$P < .026$
Wages (in \$):								
Correlation062	1,296 ^d	0	0	0	$P < .024$	$P < .022$	$P < .022$
MAVD	3,957		4,131	4,174	4,044	$P < .019$	$P < .006$	$P < .100$
Education (in years):								
Correlation111	1,482	0	0	0	$P < .001$	$P < .001$	$P < .001$
MAVD	1,772		2,013	2,093	1,881	$P < .001$	$P < .001$	$P < .007$
Gender (1 = male):								
Correlation182	1,431	0	0	0	$P < .001$	$P < .001$	$P < .001$
MAVD331		.420	.428	.406	$P < .001$	$P < .001$	$P < .001$

^a Average correlations and MAVDs generated in 1,000 samples randomly matching referrers with all applicants, nonreferrals, and referrals.

^b The lowest level of probability attainable with 1,000 replications is $P < .001$.

^c "Correlation" is the Pearson correlation between referrers' and referrals' values on each variable; MAVD is the mean of the absolute value of the difference between referrers' and referrals' values on each variable.

^d Excludes referral applicants who reported never having had a previous job.

Our strategy for testing hypothesis 2 is to compare these observed measures of homophily against “asocial” baselines, which could also produce some degree of homophily in our data. Since one would expect people who are interested in a similar employment situation to share many characteristics, the fact that applicants are more like those already hired might also produce homophily between applicants (either referrals or nonreferrals) and referrers. According to this argument, simple self-selection on the part of applicants will produce a pool that looks like those who are already employed at the firm. If this self-selection process were to account for the levels of homophily we observed, it would call into question the interpretation of the homophily as reflecting a social link between particular pairs of individuals, that “people tend to refer people like themselves.”²¹

Using the logic of bootstrapping (see Efron and Tibshirani 1993), we developed a counterfactual test for determining whether self-selection on the part of applicants could account for the level of homophily we observed. If simple attraction to the firm were to be causing the observed levels of homophily, random pairing between people who applied and referrers should often produce as much homophily as the levels we observe here. We drew 1,000 random samples (with replacement) from the pool of applicants and randomly matched them to referrers. We then calculated homophily measures between referrers’ characteristics and the randomly matched applicant for each sample. This procedure simulates how much homophily we would expect to see in our data if the specific social tie between each referrer and their referral were to have no special role in producing homophily.

The third column of table 2 shows the expected values of the homophily measures produced by this exercise. As would be expected by random pairings, for all four variables, the average of the correlations is zero. The average of MAVD measures reveals the levels of homophily one would expect based on applicant self-selection alone in the metric of each variable. For all four variables, the difference between referrers and randomly matched referrals is greater than that observed in actual referrer-referral dyads. However, for tenure and wages, these differences are quite modest: random pairings produce MAVDs that are about 5% greater than the observed values of the MAVD (1,440 vs. 1,376 days of tenure, and \$4.13 vs. \$3.96 in wages). In contrast, the MAVDs for the random baselines on

²¹ The fact is we do not directly observe referrer’s recruitment efforts, so we do not know whether “people refer others like themselves.” We only observe recruitment attempts where the referral actually applies. Strictly speaking, these data can only test the proposition that “people produce referral candidates who are like themselves.”

education and gender are considerably larger than the observed MAVD values, 14% and 27%.

While random pairings produce less homophily than that observed in these data on average, the question of statistical significance remains to be addressed. Using a bootstrapping logic (Efron and Tibshirani 1993), we test hypothesis 2 by examining how often in 1,000 trials the randomly generated levels of homophily meet or exceed the observed levels of homophily. The sixth column of table 2 reports the results of these tests. For three of the four characteristics, the randomly generated correlations *never* meet or exceed the observed correlations between referrers and referrals. For the smallest observed correlation (wages), the occurrences are rare (23 out of 1,000 samplings), rare enough to yield confidence above traditional levels of statistical significance ($P < .024$). When using the correlation as a homophily measure, hypothesis 2 is supported.

Turning to the MAVD measures, we examine how often the MAVD scores are less than or equal to the observed MAVD. Figure 2 illustrates the logic of the statistical test we developed using the data for tenure. It presents a histogram of the MAVD scores for tenure from the 1,000 random replications. Although rare, scores as low as the observed MAVD score of 1,376 days do occur in 1,000 samples where referrers are randomly matched to applicants (17 out of 1,000 draws). Therefore, the chance that the observed level of homophily on tenure (i.e., 1,376) is due to simple self-selection on the part of applicants is less than 18 out of 1,000 ($P < .018$). The P -values for the other variables in column 6 also yield comfortable levels of statistical confidence. For education and gender, random pairing of referrer with referrals never produce MAVDs that are less than equal to the observed MAVD scores in 1,000 trials. For wages—the variable showing the lowest correlation between referrers and referrals—only 18 of 1,000 random replications produce MAVD scores as low as the observed MAVD of \$3.96. Therefore, the results of the tests using the MAVD also support hypothesis 2.

While we think that the analyses just reported constitute fair tests of hypothesis 2, it occurred to us that there might also be another somewhat less naive baseline against which we might compare the observed data. In addition to attraction to this firm, referrals that come into this pool of applicants might also share a preference for informal job search, since referrals are people who, by definition, looked for work by informal means, and these applicants might also be similar to referrers who, by definition, are also involved in informal job searches. Random pairings between informally searching applicants and referrers, then, might often yield levels of homophily as high as those we observed. We drew 1,000 random samples of referral applicants (with replacement), randomly paired them with referrers, and calculated homophily measures between

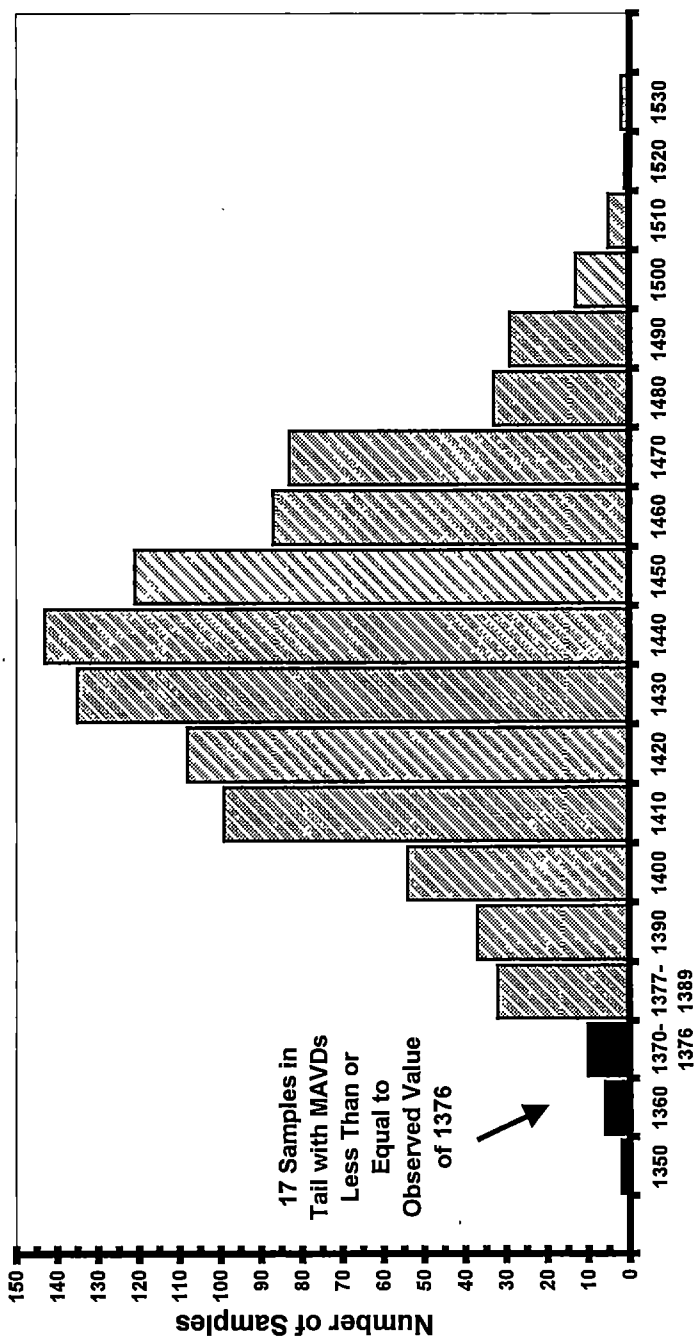


FIG. 2.—Distribution of mean absolute value of difference (MAVD) in days of tenure for 1,000 random pairings of referrers and applicants.

the characteristics of the referrer and the referral. Columns 5 and 8 of table 2 report the results of tests against this alternative baseline.

As with the baselines reported in column 3, the correlations in column 5 are all zero. As shown in column 8, the randomly generated correlations never exceed the correlations observed between referrers and referrals for two of the four characteristics. Even for the least robust of these correlations—wages—these occurrences are also rare (21 out of 1,000 samplings) and yield a comfortable level of statistical significance ($P < .022$). The results in the MAVD show that random pairings between referrers and randomly paired referrals are more similar than random pairings of referrers and applicants (see cols. 3 and 5). The statistical tests (col. 8) show that random pairings of referrers with referrals never produce MAVD scores as low as the observed values for education and gender (yielding a $P < .001$) and only once for tenure ($P < .002$). The result for wages is the least reliable but is still significant at the 10% level: 99 out of 1,000 samples yielded MAVD less than or equal to the observed MAVD ($P < .100$). We have no evidence that self-selection on the basis of informal job search (in addition to self-selection in application to the firm) can explain the degree of homophily we observe. Consequently, asocial processes of self-selection at application cannot explain our evidence supporting hypothesis 2. When considered against all the baselines (all applicants, referrals, and nonreferrals),²² people do appear to *produce* referral applications from people like themselves.

Mechanism 4: Referrals Possess Better Information of the Job

The results so far support the richer pool argument. Referrals present more appropriate applications in terms of easily measured characteristics. Although we have not been able to rule out mechanism 1 (expanded recruitment horizon) and mechanism 3 (reputation protection) processes, we have offered positive evidence in favor of mechanism 2 (homophily). We now turn to mechanism 4, the idea that the referral tie serves as an information conduit. As we discussed above, this process is a key component of at least some versions of economists' better match models of recruitment via referrals.

²² For completeness, we also performed a set of homophily tests randomly pairing referrers with nonreferrals (table 2, cols. 4 and 7). This baseline describes how much homophily we would expect from applicants who do not use informal means for job search. The statistical tests (col. 7) show that the observed levels of homophily cannot be accounted for by attraction to the firm *sans* informal job search. Column 4 shows that on average the nonreferral applicant pool is less like referrers than the pool of referral applicants; this is consistent with the idea that referrals are better at application than nonreferrals (hypothesis 1).

TABLE 3
SUMMARY OF EMPIRICAL TESTS OF THE "BETTER-MATCH" HYPOTHESIS

	Prediction	Data Source
Applicant side:		
Hypothesis 3a	Referrals have extra knowledge of full-time policy	Application forms
Hypothesis 3b	Referrals have extra knowledge of temp policy	Application forms
Hypothesis 3c	Referrals have less nonresponse on expected wage	Application forms
Hypothesis 3d	Referrals have more accurate wage expectations	Application forms
Hypothesis 4	Referrals have better-timed applications	Applicant tracking records
Hypothesis 5	Referrals have higher acceptance rate of job offers	Applicant tracking records
Hypothesis 6	Referrals have better understanding of job content	Survey of new hires
Employer side:		
Hypothesis 7	Employers use referrers' characteristics as an indirect screen	Applicant tracking records
Hypothesis 8	Employers directly contact referrers	Interviews with recruiters

Hypotheses 3–6 test whether referrals have extra information passed on to them by referrers at various stages of the screening process (see the applicant-side portion of table 3). If referrers were doing more than simply expanding the pool of formally qualified candidates (via mechanisms 1–3), then we would expect referrals' applications to show better knowledge of more tacit features of the job and the employer. Especially since referrers are paid for successful referrals, referrers have an incentive to take the time to pass on such information to referrals. Our strategy, then, is to use the application forms to identify important features of the job, which referrers would be highly likely to explain to their referrals. To the extent that information is generally available to both referrals and nonreferrals (e.g., via job advertisements), the better match model predicts that referrals have more of such information.

We develop four tests of hypothesis 3 by studying a number of important features of the firm's staffing policies with respect to the PCSR job (see hypothesis 3a–3d in table 3). We learned from our interviews with PCHR recruiters that, as a matter of policy, the firm only hires PCSRs as full-time, nontemporary employees. There is a line on the application

that asks people to check their employment preferences. The question reads: "Do you prefer" and is followed by a checklist with four boxes: "Full-Time," "Part-Time," "Temporary," and "Nights." Applicants can check all that apply. Hypothesis 3 implies that referrals should present applications that are more in line with the firm's policy. That is, referrals should be more likely to check "full-time" and less likely to check either "part-time" or "temporary."

We find no evidence of referrals having better information about hiring full-time versus part-time workers for this job. If we consider strictly incorrect answers (i.e., checking only the part-time box), referrals are not better informed than nonreferrals. Referrals and nonreferrals are equally likely (2.2%) to check only "part-time" from among "part-time" and "full-time" boxes. Virtually equal percentages of referrals and nonreferrals checked both "full-time" and "part-time" (9.6% vs. 8.4%), a difference that is in the wrong direction. The distribution of strictly correct answers (i.e., checking full-time only) is less extreme for the applicant pool overall (87.2% fall into this pattern). However, referrals and nonreferrals do not differ in their propensity to check only the "full-time" box on the application. For nonreferrals, 86.7% fall into this pattern, compared with 88.2% for referrals. Here, too, the difference is not statistically significant ($P < .163$; LR $\chi^2 = 1.942$; $df = 1$). Clearly, information on the staffing policy with respect to hours is well known by applicants and does not differ by recruitment source. We find no reliable evidence of referrals having better information of this key feature of the job; hypothesis 3a is not supported.

However, referrals were less likely than nonreferrals to check the "temporary" box on the application (.8% vs. 1.9%; $P < .005$; LR $\chi^2 = 7.734$; $df = 1$). Since the firm does not hire temps into the PCSR job, this is consistent with referrals having better information on the firm's hiring policies. However, it is hard to reconcile the idea that referrals have better information about the firm's policy on hiring temps with their apparent lack of extra knowledge of the policy of only hiring full-time employees. An alternative explanation for this pattern may be that referrals tend to come from a more stable applicant pool than nonreferrals and, therefore, are less prone to apply for temporary jobs. Although this is speculation, we do find support for this alternative explanation in our analyses of hypothesis 2 where referrals display more stable work histories. Nevertheless, by this measure, we do find support for hypothesis 3b.

We also examined referral/nonreferral differences in knowledge about another prominent feature of the PCSR job: wages. Knowledge of wages is crucial for labor economists. Indeed, no single piece of labor market information conveys more to job seekers than the wage a job pays. One model of personal networks in the labor market (Mortensen and Vishwa-

TABLE 4

TESTS OF REFERRALS DISPLAYING BETTER KNOWLEDGE THAN NONREFERRALS OF THE STARTING WAGE POLICY USING RESPONSES TO THE ITEM ON EXPECTED WAGES

	Referrals (<i>N</i>)	Nonreferrals (<i>N</i>)	Significance Test
Hypothesis 3c: referrals show less nonresponse on expected wage:			
% nonresponse	42.3 (1,481)	42.6 (2,491)	LR $\chi^2 = .031$ (<i>df</i> = 1) <i>P</i> < .861
Hypothesis 3d: referrals show more accurate wage expectations:			
% exactly \$8.25	17.5 (855) ^a	15.6 (1,431) ^a	LR $\chi^2 = 1.495$ (<i>df</i> = 1) <i>P</i> < .221
Mean absolute value of difference from \$8.25	\$0.88 (855) ^a	\$0.94 (1,431) ^a	<i>t</i> = 1.470 (<i>df</i> = 1; 2,284) <i>P</i> < .071 (one-tailed)

^a Excludes cases who did not respond to the question on expected wage.

nath 1994) is explicitly driven by the idea that job seekers get better information on offered wages through personal contacts than they do via formal means. Applicants are asked to fill in a blank for "starting salary expected." Here, too, the firm has a strict policy. With no exceptions, the PCSR job has a starting wage of \$8.25 per hour throughout the entire period of our study. If referrals have better information about the job than nonreferrals, they should provide more accurate wage expectations on the application form. We studied this issue from two perspectives. First, we examined whether referrals are any less likely than nonreferrals to have left this question blank on the application (hypothesis 3c). Second, for those who did supply a starting wage, we examine whether referrals are any more accurate than nonreferrals (hypothesis 3d).

Table 4 presents the tests of hypotheses 3c and 3d. With respect to patterns of nonresponse, we find no evidence that referrals are better informed than nonreferrals: 42.3% of referrals and 42.6% of nonreferrals chose not to fill in a starting salary. To the extent that nonresponse to this item indicates ignorance of the starting wage policy, then the data in table 4 clearly do not support the notion that referrals are better informed. Neither is there evidence supporting hypothesis 3d. If we look at the percentages responding exactly \$8.25, 17.5% of referrals versus 15.6% of nonreferrals gave the precisely correct response of \$8.25, a difference that is not

statistically significant.²³ We also tested whether referrals' starting salary responses were any more accurate than nonreferrals' responses by calculating the mean absolute value of the difference between their expected starting salary and \$8.25. Unlike the percentage giving the precisely correct response of \$8.25, this test incorporates quantitative information from throughout the distribution. On average, referrals' responses deviated from \$8.25 by \$.88, whereas nonreferrals' responses deviated \$.94. Here, too, the difference is not statistically significant. By all these measures, we find no evidence that referrals are better informed about the firm's wage policies than are nonreferrals.

As a final check of hypotheses 3*c* and 3*d*, we closely compared expected wages for referrals and nonreferrals across their entire distributions and found only minor differences distinguishing them. The median wages for referrals and nonreferrals are identical at \$8.00, and the \$.08 difference in mean wages (\$7.76 vs. \$7.84) is not statistically significant ($P < .066$, one-tailed test; $t = 1.513$). To the extent there is a difference, it appears to be due to the fact that the expected wage distribution for nonreferrals has a somewhat longer right tail than that of referrals (see fig. 3). Across most of the distribution, the plot of the percentile distribution for referrals versus nonreferrals closely tracks the 45-degree baseline in figure 3. The only consistent exception to this pattern occurs after \$9.08, the 95th percentile point for both distributions.

This latter pattern is not likely to be due to referrals being better informed than nonreferrals. If uninformed applicants—whether referrals or nonreferrals—think that reporting high past wages can influence the size of the wage offer, then formerly highly paid applicants would tend to report higher expected wages. Since more nonreferrals than referrals are drawn from jobs that pay substantially more than \$8.25 an hour (recall the pattern in fig. 1), this bargaining behavior would produce the observed result where nonreferrals show a longer right tail than referrals even if referrals and nonreferrals were equally uninformed. The data show significant support for this line of reasoning. Consistent with the idea that high wage applicants report high expected wages, there is a significant correlation (.308) between wages on last job and expected wages among applicants overall. If referrals were to be better informed about the strict starting wage policy than nonreferrals, then fewer referrals than nonreferrals would cite high expected wages, and this should result in a lower correlation between expected wages and wages on last job for referrals than nonreferrals. The pattern observed in these data, however, is the

²³ Recalculating percentages on the base of all applicants (for respondents and nonrespondents) does not change this conclusion: 10.1% of referrals and 9.0% of nonreferrals fill in precisely correct responses for starting wage ($P < .221$; LR $\chi^2 = 1.496$; $df = 1$).

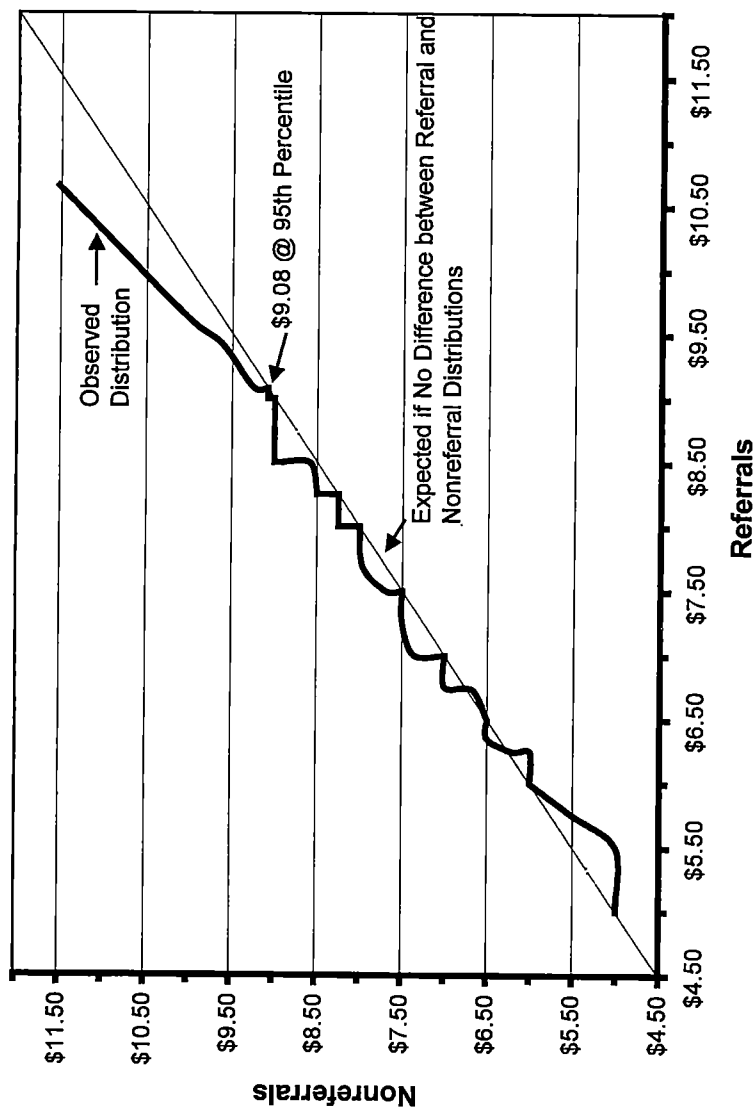


FIG. 3.—Percentile distribution of expected wages by referral status

exact opposite of this: the correlation between wage on last job and expected wage among referrals is *higher* than it is for nonreferrals (.326 vs. .298). Although this difference is not statistically significant,²⁴ the fact that the direction of the contrast is in the wrong direction casts serious doubt on the idea that referrals' shorter right tail is due to referrals' better information regarding PCHR's policy on starting wages. More plausibly, it is the fact that nonreferral applicants are drawn from jobs with higher wages than that of referrals that accounts for their displaying a longer right tail in figure 3.

In hypothesis 4, we conjectured that, if referrals were being passed systematically better information about the company's hiring needs, they might use this information to time their applications such that they would be applying when it would be comparatively easier to be successful and be hired (Fernandez and Weinberg 1997). Especially in light of referrers' financial stake in having their referrals make it through the screening process, we might expect referrers to pass on this sort of information. We tested this proposition by comparing the average state of the market for referrals and nonreferrals. For each application, we coded the number of job openings and applications on the date the person applied. We found no support for this process. Referrals applied on days when there were *fewer* job openings than did nonreferrals (17.4 vs. 18.6). Referrals did apply on days when there were slightly fewer applicants than nonreferrals (19.5 vs. 20.2), but the difference was not statistically significant ($P < .109$, one-tailed test; $t = 1.213$). Nor was there evidence of referrals better timing their applications when we examined the ratio of the two variables. The difference was not in the predicted direction: applicants per opening averaged 1.83 on days that referrals applied versus 1.75 on days that nonreferrals applied. We also experimented with various leads (where referrals might be applying in anticipation of openings) and lags (the number of job openings the previous week and month). Similar to Fernandez and Weinberg (1997), we found no support for hypothesis 4 using these alternative formulations. Overall, hypothesis 4 is not supported.

Thus far, we have examined whether referrals possess superior information about various facets of the job than nonreferrals using data from the application stage of recruitment. The last two tests of the better match

²⁴ In OLS regressions predicting expected wage, we tested for an interaction between referral status and wage on the last job and found that it was insignificant. Indeed, referral status is not significantly related to expected wage, either individually or in concert with wage on last job. In models including only the dummy variable for referral, the t -value for referral is -1.513 ; in models with the main effects of last wage and referral, the t -value is $-.494$. The final model, which includes the interaction between referral and last wage, yields a t -value of .608 for the interaction and a t -value of $-.743$ for the referral dummy.

TABLE 5
MEAN RESPONSE ON JOB INFORMATION ITEM FOR REFERRED AND NONREFERRED
HIRES

	REFERRALS			NONREFERRALS	TOTAL
	Job Explained by Referrer	Job Not Explained	All Referrals		
Mean response ^a	4.41	4.07	4.31	4.28	4.30
No. of cases	96	40	136	97	233

^a Responses to the questionnaire item: "Rate your understanding of what the position responsibilities and job content would be prior to accepting this position" (from 5 = very good understanding to 1 = very poor understanding).

mechanism involve later stages in the screening process. Another inference that flows from referrals' being better informed than nonreferrals is that referrals should be presold on the job and hesitate less than nonreferrals in accepting job offers (Ullman 1966). Consequently, referrals should be more likely to accept job offers once they have been extended. Given a job offer, the acceptance rate for referrals was 94.0%, compared with 90.1% for nonreferrals. This difference is not statistically significant ($P < .175$; LR $\chi^2 = 1.839$; $df = 1$). Here, too, we do not find statistically reliable support for hypothesis 5. We found a similar pattern in a retail bank: 90% of both referrals and nonreferrals accept job offers (Fernandez and Weinberg 1997).

As a final test of the applicant's side of the better match story, we look for evidence of referrals' extra information among hires. During training sessions at the beginning of their second week of employment, PCHR collected anonymous survey data from new hires. Both referrals and nonreferrals were asked: "Rate your understanding of what the position responsibilities and job content would be prior to accepting this position." The answers were recorded on a five-point Likert scale (from 5 = very good understanding to 1 = very poor understanding). Prior to this question, respondents were asked whether they were a referral ("Were you referred by a [NAME OF FIRM] employee?"). People identifying themselves as referrals were also asked, "Was this job explained to you by this person?" Thus, these unique data allow us to develop a relatively direct test of the mechanism that referrers are passing "extra" job information on to referrals during screening.

Table 5 shows that the clear majority of referrals (70.1%) report having had the job explained to them by the referrer. As might be expected, having the job explained by the referrer results in respondents reporting a higher level of understanding of the job: mean of 4.41 for referrals who

had the job explained versus 4.07 for referrals who did not have the job explained.²⁵ However, a sizeable percentage of referrals (29.2%) are being hired without the benefit of referrers explaining the job to them, and non-referrals also report very high levels of understanding of the job. Indeed, nonreferrals' level of understanding is virtually identical to that of the population of all referrals (4.28 vs. 4.31, a difference that is not statistically significant). This is contrary to the better match theory's prediction that referrals on average will have extra information vis-à-vis nonreferrals. Consequently, hypothesis 6 is not supported.²⁶

Mechanism 4: The Employer Possesses Better Information on Referrals

Hypotheses 7 and 8 address the employers' side of the better match theory. We start by considering the evidence with respect to the indirect means by which employers may harvest information about referrals. While our findings with respect to hypothesis 2 (homophily) suggest that information on applicants is available from considering referrers' characteristics, we do not know if recruiters are actually using this information. In order to test hypothesis 7, we look for evidence that recruiters are using "upstream" information (i.e., information garnered from the characteristics of referrers) when making screening decisions (see the employer-side section of table 3).

Among referral applicants, 64.8% are granted interviews with hiring managers compared to 57.5% of nonreferrals, a difference that is statistically very reliable ($P < .0001$; LR $\chi^2 = 21.905$; $df = 1$). Referrals also

²⁵ This is the only contrast in table 5 that is statistically reliable ($P < .007$; t -test of 2.72; $df = 1,133$).

²⁶ We should make one final point with respect to our tests of the applicants' side of the better match hypotheses (hypotheses 3–6). All of the people who applied to this firm are self-selected in the sense that they had enough interest in the firm to pursue a job there. However, to the extent that referrals have jobs explained to them by referrers, and decide not to apply for the job, then the survivors among the referrals may be more selected than the nonreferrals. Differential selectivity is predicated on the idea that, compared with nonreferrals, referrals have extra knowledge of the job given to them by their referrers. The fact that superior information can be used as a basis for deciding not to apply for a job at the firm suggests that there is a possibility of selection bias in our assessment of the better match theory (Winship and Mare 1992). In this case, however, selection bias cannot explain this pattern of results. To the extent that survivors among the referrals are more selected than the nonreferrals, it is referrals with extra knowledge of the job that will be overrepresented at later stages of the hiring process. Consequently, differential selection in this case would work in the direction of finding differences between referrals and nonreferrals on knowledge of the various job features we study. To have found such little evidence of extra information in a context that is biased in the direction of finding such differences adds to our confidence that the mechanism 4, "referrals have superior information," process is not at work in this environment.

receive job offers at much higher rates than nonreferrals. Conditional on interview, referrals are offered jobs at a rate that is over 1.5 times the nonreferrals' rate (18.3% vs. 11.6%). Here, too, this difference is highly significant ($P < .0001$; LR $\chi^2 = 21.584$; $df = 1$). Referrals' advantages at each stage compound. When the job-offer rate is calculated on the base of all applicants (as opposed to interviewees), 11.9% of referrals receive job offers compared with only 6.7% of nonreferrals ($P < .0001$; LR $\chi^2 = 32.180$; $df = 1$).

While referral candidates clearly enjoy advantages over nonreferrals at each stage in the hiring process, these advantages do not necessarily imply that recruiters are using information garnered from the characteristics of referrers when making screening decisions. These differences could simply be due to referrals being more appropriate applicants (hypothesis 1). Before concluding that referrals are using "upstream" information on the referrer during screening, we examine whether recruiters use these referrers' characteristics in screening. Our strategy for testing hypothesis 7 is to model recruiters' screening decisions and to see whether adding referrers' characteristics significantly improves the fit of the model.

We divide these analyses to reflect the two phases of screening: interview and offer. Table 6 presents descriptive statistics for the interview and job-offer models. For both models, we include the nine factors for which PCHR says they screen as predictors. Here, too, we coded tenure on the last job and wage on the last job as zero for people who had not had a previous job, so that the tenure and wages effects are conditional on the number of previous jobs being one or greater. However, we also add a number of variables to the model in order to replicate as closely as possible our analyses of hiring in another unit of the bank (Fernandez and Weinberg 1997). Because the unit of analysis is the application, and some people applied multiple times, we include a dummy variable to distinguish repeat applicants from first-time applicants (1 for repeat applicants, 0 otherwise; for a similar procedure, see Fernandez and Weinberg [1997]). The maximum number of applications from individuals is three. Of the original 4,165 employment inquiries, 388 (9.3%) were from individuals who had applied twice, and only 12 (0.3%) applied three times. Work in the human capital tradition (e.g., Mincer 1974) argues that the value of work experience declines over time. We capture this effect by entering a squared term for months of nonbank experience.²⁷ Although they did

²⁷ In preliminary analyses, we examined a number of specifications of the experience variables. In particular, we tested whether there was evidence of diminishing returns on the various experience measures. We found no evidence of diminishing returns to months of banking or customer service experience. In all these tests, the t -values on the relevant coefficient were always less than one. We present the model with linear

TABLE 6

MEANS AND STANDARD DEVIATIONS FOR VARIABLES IN THE INTERVIEW AND JOB-OFFER MODELS

	INTERVIEW MODEL		JOB-OFFER MODEL	
	Mean	SD	Mean	SD
Independent variables:				
Gender (1 = male)334	.472	.311	.463
Repeat application (1 = yes)096	.294	.093	.290
Skills:				
Computer788	.409	.823	.382
Language201	.401	.188	.391
Years of education	13.818	1.800	13.949	1.777
Experience:				
Months of bank experience	2.110	14.752	2.521	16.183
Months of nonbank experience	65.638	59.026	70.723	61.630
Nonbank experience, squared	7,791.197	17,391.139	8,798.011	19,187.734
Months of customer service	33.533	45.086	37.192	48.721
No. of previous jobs	3.279	1.046	3.296	1.053
Works at time of application528	.499	.562	.490
Tenure in last job (in days)	706.193	1,051.016	787.682	1,145.258
Wage in last job	7.368	3.047	7.387	2.895
Application behavior:				
No. of applications	18.992	15.537	18.510	14.797
No. of job openings	19.131	10.972	21.125	11.851
Application source:				
External referral352	.478	.381	.486
Referrer's characteristics at time of referral application:				
Tenure	1.520	3.130	1.640	3.192
Wage	3.643	5.718	3.879	5.574
Years of education	4.386	6.001	4.763	6.120
Dependent variables:				
Interviewed617	.486
Received job offer136	.343
No. of cases	2,987		1,843	

not say that they screen for these variables in this setting, we also add dummy variables for evidence of foreign language skills (applications had a line asking for such information) and gender (1 = male) as controls. In addition, we also control for the state of the market in these analyses, that is, the number of job openings and the number of applications on the date the candidate applied.

effects of bank, nonbank, and customer service experience, and a squared effect for nonbank experience because it is the best fitting specification.

Table 7 shows results of probit models predicting interview among applicants.²⁸ The first model reports the effects of the various background factors without controlling for referrals versus nonreferrals. Considering first the work experiences variables, recruiters are more likely to grant interviews to applicants who are employed at the time of applications. Candidates reporting more months of customer service experience and work experience outside the financial services sector are also more likely to be granted interviews, although the squared term on the nonbank experience variable shows that there are diminishing returns to nonbank experience. Candidates reporting longer tenure on their last job are also more likely to be interviewed.

However, two of the variables measuring work history did not emerge as significant predictors, although their coefficients had the expected sign. The number of previous jobs that candidates report on their applications is not significantly related to being granted an interview. Nor is experience in the financial services sector a significant predictor of being interviewed. This is not due to multicollinearity among the various experience measures. Even when the other experience measures are removed, bank experience is not significantly related to interview. Inspection of the mean for months of bank experience in table 6 shows the likely reason why this is the case. The average number of months of bank experience in this pool of applicants is only two months, compared with over five years for nonbank experience and more than two and half years of customer service experience.

We also find that more highly educated candidates and applicants reporting computer skills are more likely to be interviewed.²⁹ Applicants with foreign language skills, however, are *less* likely to be interviewed than applicants without such skills. While we can only speculate, this effect could be a by-product of recruiters' screening on verbal and "soft" interpersonal skills, which they glean from phone calls or short interviews. As one recruiter put it: "They [candidates] have to speak English." Because all customer interactions are conducted over the telephone, PCSRs

²⁸ In preliminary analyses, we tested whether the models predicting interview and job offer (tables 7 and 8) differ for referrals and nonreferrals. We found no evidence of statistically significant interactions with referral status for these models.

²⁹ In preliminary analyses, we broke the education variable into a series of splines (less than 9 years, 9–12 years, 13–16 years, and more than 16 years) in order to test whether very highly educated applicants are being treated as "overqualified" and are less likely to be interviewed. Although the spline for applicants with over 16 years of education shows that the probability of interview is lower than for other candidates, this effect is not statistically reliable (the other splines show very linear increases in the probability of interview). This is probably due to there only being 122 such cases.

TABLE 7
COEFFICIENTS FOR THE PROBIT REGRESSION PREDICTING JOB INTERVIEW FOR CUSTOMER SERVICE REPRESENTATIVE JOB ON SELECTED
INDEPENDENT VARIABLES

INDEPENDENT VARIABLES	MODEL 1			MODEL 2			MODEL 3		
	Coefficient	SE		Coefficient	SE		Coefficient	SE	
Constant	-1.419***	.208		1.547***	.211		-1.559***	.211	
Gender (1 = male)	-.150**	.053		-.145***	.053		-.147**	.053	
Repeat application (1 = yes)	-.023	.082		-.035	.082		-.040	.083	
Skills:									
Computer292***	.060		.289***	.060		.290***	.061	
Language	-.163**	.061		-.173**	.061		-.176**	.061	
Years of education063***	.015		.067***	.015		.067***	.015	
Experience:									
Bank experience001	.002		.002	.002		.002	.002	
Nonbank experience004***	.001		.004***	.001		.004***	.001	
Nonbank experience, squared	-.000009**	.000003		-.000009**	.000003		-.000009**	.000003	

Customer service experience002**	.001	.002**	.001	.001
No. of previous jobs	-.009	.026	-.007	.026	.026
Works at time of application159***	.049	.128**	.049	.049
Tenure on last job000101***	.000032	.000095**	.000032	.000032
Wage on last job	-.028***	.009	-.027**	.009	.009
Application behavior:					
No. of applications	-.000079	.001561	.000082	.001565	.001566
No. of job openings030***	.002	.031***	.002	.002
Application source:					
External referral223***	.052	.429
Referrer's characteristics at time of referral application:					
Tenure012
Wage010
Years of education035
χ^2 (df)	324.717*** (15)		343.088*** (16)	349.185*** (19)	
Improvement χ^2 (vs. previous model)			18.371*** (1)	6.097 (3)	

NOTE.—No. of cases = 2,987.

* $P < .05$.

** $P < .01$.

*** $P < .001$, two-tailed test.

cannot have heavy accents; this foreign language effect could be picking up recruiters' screening on such phone skills. The final human capital variable we consider is candidate's last salary on the job. Controlling other things, we find applicants who report a higher wage on their last job as being *less* likely to be granted interviews. As we mentioned above, recruiters worry that such candidates might be overqualified and more likely to leave.³⁰

If we consider the two controls, repeat applicants are no more likely to be interviewed than are first-time applicants. However, controlling other factors, males are less likely to be interviewed than females. PCHR recruiters speculate that females have a better sense of how to conduct customer service interactions and that this may come across in screening. Females do report significantly more months of customer service experience than do males (37.1 vs. 28.5). However, we could not find any statistically reliable evidence of interactions between gender and other variables in preliminary analysis.

The second model in table 7 adds a dummy variable distinguishing between referrals and nonreferrals. The probit coefficient for referral is statistically significant and remains about the same magnitude as the coefficient with no controls (point estimates of .223 with controls vs. .212 without controls). Although referrals appear to be more appropriate candidates for the PCSR job (see the evidence with respect to hypothesis 1), referrals' advantages at the interview stage cannot be explained by the individual background control variables. Moreover, the introduction of the referral variable does not change the pattern of the other effects (see models 1 and 2), indicating that whatever is leading recruiters to prefer referrals at this stage, it is relatively independent of the background factors. While it is plausible that we would find this pattern if recruiters were using referrer's "upstream" information in their screening decisions, this pattern is also consistent with the notion that referrals are more desirable than nonreferrals in other unmeasured ways.

The third model in table 7 seeks to provide evidence of recruiters using upstream information in their recruiting by adding three key referrers' characteristics to the equation. For each referral, we coded referrers' wage, tenure with the company (in years), and years of education, all measured at the time the referral was made. For nonreferrals, these variables are coded as zero. Hence, the effects of referrers' characteristics are conditional on the applicant being a referral. In the better match story, these variables measure different aspects of referrers' quality as an employee. Wage is a measure of the referrer's value to the firm, whereas longer ten-

³⁰ As with the experience and education effects, we experimented with various nonlinear specifications and found that the simple linear effect of last wage is the best fitting.

ure should be associated with lower turnover, and referrer's education is a measure of the referrer's human capital. The chi-square comparing model 3 with the model with just the dummy variable (model 2) shows that these three measures do not significantly improve the fit of the interview model ($P < .107$; LR $\chi^2 = 6.097$; $df = 3$). Moreover, the coefficient on referrer's wage is in the wrong direction: referrals from high-wage referrers are *less* likely to be granted interviews than referrals from low-wage referrers. At the interview phase, hypothesis 7 is not supported.³¹

Table 8 reports the results of a similar analysis predicting job offer among candidates who received interviews.³² Across all three columns,

³¹ Another approach to testing hypothesis 7 is to study only referrals (i.e., excluding the nonreferrals) and to add referrers' characteristics to a baseline model that has only individual background variables. The results using this method are identical to those we present here. We prefer the models including nonreferrals because they have increased statistical power over models using only referrals (recall that we found no evidence that the effects are different for referrals vs. nonreferrals); the models distinguishing referrals and nonreferrals with a dummy variable (i.e., model 2 of tables 7 and 8) will become important later when we discuss a version of the better match process (see below).

³² The fact that the cases analyzed in table 8 are all survivors of the interview stage introduces the possibility that selection bias may affect our assessment of hypothesis 7 at the job offer stage. In preliminary analyses, we attempted to control for such selection bias using a bivariate probit model with selection, which is the appropriate statistical procedure when both the ultimate dependent variable (job offer) and the selection criterion (interview) are dichotomous. In contrast with a study of the western region (Fernandez and Weinberg 1997), we were less than wholly successful in this case. When working with the full model where all of the variables in column 3 of tables 7 and 8 are included in both the interview and job offer equations, we could not get the model to converge. Such estimation problems are common with this model, since it is only weakly identified (off of the nonlinearity of the selection effect). In order to obtain estimates, we needed to define "instruments"—variables that, by assumption, affect the selection stage but not the substantive stage. We were able to obtain estimates if we dropped the number of jobs and number of applicant's variables from the job offer equation (but not the interview equation). This is tantamount to arguing that PCHR recruiters worry about the state of the market when deciding whom to interview but that line managers have delegated concerns about the state of the market to PCHR when deciding job offers. Under this model of the process, the results from the two phases look very similar to one another, and we find no evidence of selection bias accounting for our findings with respect to hypothesis 7. We need to point out that the hiring process at the phone center is organized in such a way to make it very difficult to identify selection bias. To the extent that PCHR recruiters are successful in mimicking the behavior of the hiring manager, PCHR's actions become indistinguishable from those of the hiring manager. In the limit, one can consider them becoming hiring managers. To the extent the stages meld into one, it makes no sense to control for selection. This would be like trying to study the predictors of some dependent variable, after controlling for the same dependent variable. The fact that PCHR recruiters were granted such hiring authority after our field period suggests that the hiring process we study might have been approaching this limit.

TABLE 8
COEFFICIENTS FOR THE PROBIT REGRESSION PREDICTING JOB OFFER FOR CUSTOMER SERVICE REPRESENTATIVE JOB ON SELECTED
INDEPENDENT VARIABLES

INDEPENDENT VARIABLES	MODEL 1		MODEL 2		MODEL 3	
	Coefficient	SE	Coefficient	SE	Coefficient	SE
Constant	-.976**	.342	-1.182***	.348	-1.145***	.350
Gender (1 = male)	-.248**	.088	-.247**	.089	-.255**	.090
Repeat application (1 = yes)	-.154	.137	-.170	.138	-.156	.138
Skills:						
Computer146	.103	.162	.104	.150	.105
Language	-.071	.101	-.090	.102	-.097	.102
Years of education	-.007	.023	-.003	.023	-.006	.023
Experience:						
Bank experience000032	.002346	.000022	.002346	.000046	.002339
Nonbank experience004**	.002	.004**	.002	.005**	.002
Nonbank experience, squared	-.000012*	.000005	-.000012*	.000005	-.000012*	.000005

Customer service experience001	.001	.001	.001	.001
No. of previous jobs	-.055	-.048	.039	-.051	.040
Works at time of application327***	.280***	.080	.295***	.081
Tenure on last job000051	.000049	.000041	.000050	.000041
Wage on last job	-1.074	-.011	.016	-.010	.016
Application behavior:					
No. of applications	-.007*	-.007*	.003	-.006*	.003
No. of job openings	-.009**	-.009**	.003	-.009**	.003
Application source:					
External referral317***	.077	-.258	.532
Referrer's characteristics at time of referral application:					
Tenure				-.036*	.017
Wage008	.014
Years of education051	.043
χ^2 (df)	69.619*** (15)	86.548*** (16)		93.759*** (19)	
Improvement χ^2 (vs. previous model)		16.929*** (1)		7.211 (3)	

Note.—No. of cases = 1,843.

* $P < .05$.

** $P < .01$.

*** $P < .001$, two-tailed test.

none of the skill measures emerge as statistically reliable predictors of job offer. Among the experience variables, only nonbank experience, its squared term, and the dummy variable for working at the time of application remain as statistically significant predictors. The pattern of effects for the application behavior variables changes compared to the interview equation. Conditional on interview, both the number of applications and the number of job openings on the day the candidate applied are negatively associated with being offered a job. We do not see a substantive explanation for why this should be the case. However, when paired with our experiences trying to control for selection bias (see n. 32), we suspect that these sign reversals are due to the fact that all the cases in the offer equation have been interviewed.

The second model in table 8 shows the same pattern as table 7: controlling other factors, line managers are more likely to offer jobs to referrals. However, similar to the model predicting interview, including referrer's characteristics does not improve the fit of the model. Moreover, the effect of the referrer's tenure variable is in the wrong direction. Here, too, hypothesis 7 is not supported. To the extent we have been able to control for its influence, this finding is robust to selection bias (see n. 32).

Employers can also learn more about referrals than nonreferrals by directly contacting referrers and asking them about the referral. We addressed hypothesis 8 by interviewing the PCHR personnel responsible for recruiting for the PCSR job. After completing the other analyses for this article, we asked the people whose job it would be to contact referrals questions about how they screened referrals.³³ We asked the following direct question: "When an application from a referral comes across your desk, do you contact the person who referred them?" The PCHR personnel involved in screening for the PCSR job all answered unequivocally, "No, never." We next asked: "Do you look up any of the characteristics of the referring person—such as their tenure with the company or their salary—in your HR database?" The answer was again an emphatic, "No, never." These responses are clearly contrary to the predictions of the better match story. Hypothesis 8 is clearly not supported. Moreover, the fact that recruiters never contact referrers also explains why we saw no support for hypothesis 7.

³³ Prior to this point, PCHR personnel's answers to our questions about how they recruited had been that they "treated referrals the same as everyone else." Since we knew referrals were being interviewed and hired at higher rates than nonreferrals (see above), we were concerned that these responses revealed more about their ideology than their actions. Only after completing the analysis of their interviewing and offering behavior did we feel that we could usefully probe them on the specifics of what they do.

We asked these recruiters to elaborate on why they did not contact the referrer or look up information on them. While PCHR personnel understood the ideas of “upstream information” and that “people tend to refer people like themselves,” they explained that they simply do not have the time to do these things. In failing to invest time in these relatively easy procedures when hiring referrals, we suggest that the PCHR department is revealing in its actions that it does not place much stock in the better match theory. Indeed, one individual recruiter’s impression was that referrals were *less* well matched than nonreferrals, post hire. She explained, “Referrals are worse; I think they have higher turnover rates than people we get from a newspaper ad. You would think that people would not refer just anybody since it would reflect on them if the person were not any good. But the other side of this is that I know people who would refer their dog if they can get a \$250 bonus.” While this recruiter shows a clear understanding of the reputation protection argument in this quote, she also expresses skepticism that the \$250 referral bonus is likely to decrease turnover via the better match mechanism. We heard similar sentiments expressed by other human resources personnel across a number of different units of the bank.³⁴

Although we did not raise these to the level of hypotheses, we also checked whether there is evidence of the various posthire predictions of the better match theory. More than just idle curiosity motivated us to perform these checks. As we mentioned above (see n. 9), some versions of the better match story argue that better matches could occur for referrals *without* the referrer passing on any explicit information to the referral or the employer (e.g., Simon and Warner [1992] make no reference to referrers passing on information in his model). According to these arguments, all that is required for better matches is that there be homophily between the referrer and the referral (mechanism 2) and that referral applicants be more likely to apply (mechanism 1). These models, in essence, claim that, since referrers bring in people with more desirable hard-to-observe characteristics, the richer pool process should be enough to yield better

³⁴ According to these people, referral bonuses are counterproductive since people who are referring others for money will not refer the best people. In essence, these recruiters are saying that bonuses are oversufficient justification (Lepper and Greene 1978) for referring others, that bonuses replace what should be the intrinsic rewards of helping to recruit for the bank (e.g., a sense of participation in the running of the bank) with extrinsic rewards. As a recruiter in another unit put it: “[Referrers] should be doing it for the company. If they are only doing it for the money, [the referral] can’t be good for us. And the thing with bonuses is that it is a slippery slope. Before you know it, you will find yourself having to offer people bonuses just to do their job. If we continue down this path, before long we are going to have to offer bonuses for people picking up the phone to call a client.”

posthire outcomes for referrals. In the face of the evidence we have found for referrer-referral homophily (hypothesis 2), and evidence that referrals constitute a richer pool at application in terms of easily observed characteristics (hypothesis 1), posthire evidence of better matches would support this "mechanism 1 + mechanism 2" version of the better match theory. This would suggest that referrer's search produces not only applicants who are more appropriate on formal criteria, but also more appropriate on the more tacit, hard-to-screen-for characteristics. Recruiters would then use referral status as a signal of referral's posthire superiority. According to this interpretation, the effects of the dummy variables for referral status in tables 7 and 8 would be due to recruiters using referral status as a signal of these hard-to-screen-for characteristics.

On the other hand, if referrals are not better matched than nonreferrals, post hire, then there would be no signal value in referral status. Such a pattern would support the notion that what referrers are doing is delivering applicants with characteristics that are more appropriate for the job on formal criteria for which it is relatively easy to screen. Not finding evidence of referrals being better matched post hire would lead us to conclude that the "referral status as signal" version of the matching account is also lacking in this setting. In this case, the effects of the dummy variables for referral status in tables 7 and 8 should not be interpreted as evidence of a signaling process, since there is no posthire value in the signal. In this scenario, the dummy variable effects of referral status would reflect the absence of variables that we as analysts have not been able to control in the model but that recruiters do notice and use in their screening decisions.

We inspected the data from the phone center to see if we could find any evidence of these posthire implications. If referrals are better matched to their jobs than nonreferrals, then referrals should have higher starting wages (Corcoran et al. 1980; Simon and Warner 1992), slower wage growth (Simon and Warner 1992), lower turnover than nonreferrals (Corcoran et al. 1980; Datcher 1983; Simon and Warner 1992; Sicilian 1995), and flatter time paths of turnover than nonreferrals (Ullman 1966).³⁵ The evidence on wages contradicts the better match account. For *all* new hires

³⁵ It has been argued that turnover in general should show a pattern of time dependence. Bad matches should dissolve quickly, leaving only better matches intact as turnover unfolds over time. Consequently, survivors should be more likely to stay on the job the longer they are on the job (Tuma 1976; Jovanovic 1979; Lane and Parkin 1998). The matching story predicts that turnover chances should decline more precipitously for nonreferrals than referrals over time (Ullman 1966). This is because the extra information that referrals and employers exchange should lead to fewer early bad matches, which dissolve among referrals and not among nonreferrals.

into the PCSR job, the starting wages are \$8.25 per hour. This would suggest there is little reason to expect patterns of wage growth to differ and, indeed, the patterns of wage growth for referrals and nonreferrals are statistically indistinguishable.³⁶ Nor do we find statistically reliable evidence of referrals having a lower turnover rate. We performed several test statistics (i.e., Log-Rank and Wilcoxon tests) to compare the Kaplan-Meier estimates of the survivor function for referrals versus nonreferrals (see the first column of table 9). Here, too, the differences between referrals and nonreferrals are never statistically significant ($P < .1292$).³⁷

We also tested for differences in turnover rates between referrals and nonreferrals after controlling for a number of covariates (see col. 2 of table 11). Results do not change across a variety of parametric transition rate models (the Cox model, the proportional exponential model, or the proportional Weibull model). Lastly, we examined the time path of turnover hypothesis. Using parametric models of time dependence (the nonproportional Weibull models, and various piecewise-Weibull and piecewise-exponential models using one-month intervals to examine nonmonotonic patterns), we never found evidence that the (always insignificant) effect of referral changes over time. In models using the covariates listed in column 2 of table 11, referrals do not differ from nonreferrals in their time paths of turnover. *None* of the posthire predictions of the better match story are borne out in this setting.

Skeptics might question whether the sample of hires at the phone center provides sufficient power to deliver fair tests of the posthire predictions of the better match story. We addressed this issue by replicating the wage, wage growth, turnover, and time path of turnover analyses across a number of sites. We collected posthire data on 936 hires (293 referrals and 643 nonreferrals; see second column of table 9) from three other phone centers located in different cities. We performed these replications both by pooling the data across the three sites and then again separately by site. Where possible, we repeated these analyses using data on hires from a very different setting in the bank, four entry-level jobs in the western region of the retail bank.

³⁶ In analyses not presented here, we estimated wage growth equations in order to test whether referrals and nonreferrals have different patterns of wage growth. We estimated growth models separately for censored and uncensored cases and then with a Heckman (1979) selection correction for being censored versus not censored (see Winship and Mare 1992). None of these analyses show statistically reliable evidence of referrals showing lower wage growth.

³⁷ We also performed these tests separately for voluntary (quitting) vs. involuntary (fired or laid off) turnover and found no reliable differences. Most of the turnover was voluntary: only 18 (11%) of the overall job terminations were involuntary.

TABLE 9

SURVIVAL RATES OF REFERRALS AND NONREFERRALS AT FOUR PHONE CENTERS

	City 1 ^a	Total Cities 2-4	City 2	City 3	City 4
1 month:					
Referrals	97.06	99.66	98.86	100.00	100.00
Nonreferrals	98.06	99.53	97.67	99.63	100.00
3 months:					
Referrals	93.53	99.32	97.73	100.00	100.00
Nonreferrals	90.97	98.60	95.35	98.51	99.65
6 months:					
Referrals	86.47	97.61	95.45	100.00	97.89
Nonreferrals	81.94	95.80	90.70	96.64	96.54
9 months:					
Referrals	74.71	93.79	90.65	98.41	93.66
Nonreferrals	74.19	93.28	90.70	93.28	94.09
1 year:					
Referrals	70.57	90.32	84.64	95.24	91.36
Nonreferrals	66.39	90.55	89.16	89.55	92.16
18 months:					
Referrals	66.09	83.94	72.02	87.00	87.95
Nonreferrals	55.87	88.12	82.74	86.46	90.93
2 years:					
Referrals	53.71	70.67	72.02	72.78	73.42
Nonreferrals	47.82	77.92	75.85	77.98	79.16
<i>N</i> (referrals)	170	293	88	63	142
<i>N</i> (nonreferrals)	155	643	86	268	289
χ^2 (<i>df</i> = 1)	2.30	2.21	.64	.67	2.47
<i>P</i> <1292	.1375	.4233	.4119	.1160

NOTE.—Data is uncensored at sites 1 to 4 until 347, 242, 410, and 263 days, respectively. The censored data extends to 1,104, 1,690, 1,545, 1,670 days, respectively.

^a City 1 is the Midwestern site.

A lack of statistical power does not appear to account for our findings. All three phone centers had rules regarding starting pay being the same for all new hires, thus the simple wage predictions are not supported in any of those sites. Even in the western region where starting wage policies are more flexible, we did not find statistically reliable differences between referrals' and nonreferrals' starting wages. Across all replications of the wage growth models, we found no statistically reliable support for the wage growth predictions of the matching theory. Nor did we find support for the turnover predictions of the matching story (see cols. 2-5 in table

9).³⁸ This was true both with and without a number of covariates.³⁹ Our replication of the various parametric models of time dependence (the piecewise-Weibull and piecewise-exponential models) in the alternative sites does not conform to the theory's predictions either. In sum, in these other settings as well, we find no support for the posthire predictions of the matching story. At least with respect to posthire outcomes, the "referral as signal" of posthire turnover version of the matching story does not appear to be operating in this company.

While referral status does not appear to act as a signal of posthire superiority, it is still possible that referral status is being used as a signal of hard-to-observe quality factors in a prehire match story.⁴⁰ As illustrated in figure 4 with the turnover outcome, while posthire differences between referrals and nonreferrals might be nil, this could be because the prehire screening process does a good job of winnowing out bad matches, while passing on good matches who survive to be hired. Thus, it might be that referrals are better matched than nonreferrals, but at an early phase of the hiring process.⁴¹

³⁸ In these replications, we repeated the different specifications of the transition rate models, i.e., the Cox, the exponential, and Weibull models. We also performed tests separately for voluntary versus involuntary turnover and found no significant differences by referral status. As with the midwestern site, most of the turnover was voluntary: the corresponding numbers of fires are 5, 16, and 10 for each of the phone center sites. In the retail bank as well, we did not find differences in voluntary, involuntary, or overall turnover by referral status using any of the different specifications of the transition rate models.

³⁹ For the alternative sites, we did not have access to all the variables we collected in the midwestern site. The models in the alternative sites only included covariates for gender, minority versus nonminority, age, marital status, and education.

⁴⁰ We thank an anonymous reviewer for pointing this out.

⁴¹ The argument here is that prehire screening *substitutes* for the posthire turnover matching mechanism. An implication of this story is that referrals' superiority would manifest itself post hire if we were to eliminate various phases of the screening process. This can be demonstrated by moving the y-axis in fig. 4 to the left: the further back one pushes the axis, the more apparent referrals' advantage over nonreferrals would become in terms of lower turnover. Referrals' superiority would be at its maximum in the limit where hires were being made from each applicant pool on a random basis. Therefore, selection bias introduced by the screening process explains why referrals do not look any better than nonreferrals in posthire analyses. Ostensibly, running selection-corrected models would generate estimates of what referral/nonreferral differences in turnover (or wage growth) would be if we were to eliminate all prehire screening. We found no evidence for the substitution argument in preliminary analyses. The posthire results we reported above do not change after controlling for selection biases introduced by hiring. For turnover, we estimated a bivariate probit model with selection where the ultimate dependent variable was uncensored turnover (i.e., up to 347 days past hire), and the first stage was hired versus not hired using the factors listed in table 6. For the wage growth model, we corrected for selection (hired

While referrals are clearly preferred over nonreferrals in prehire screening, this could also be due to referrals having more appropriate screenable characteristics than nonreferrals (i.e., the richer pool of hypothesis 1) and would also yield the pattern depicted in figure 4. The crucial issue distinguishing the richer pool and the prehire better match accounts is the observability of the screening criteria. If the characteristics being screened for are easy to measure, then there is no independent benefit to the company in preferring referrals once recruiters apply their screen. In this scenario, knowing whether an applicant is a referral or nonreferral adds no new information once applicants have been screened. But if referrals' advantages go beyond screenable characteristics, then the prehire better match argument would predict that the company would get some additional benefit from preferring referrals, even after recruiters have applied their screen. Referrals would have more of the hard-to-measure factors that impress recruiters and make them more hireable than nonreferrals. If a new technology were to come along and make these formerly hard-to-measure characteristics easily and cheaply observable, recruiters would screen on those criteria instead of using referral status as a signal. If referrals have more of these factors than nonreferrals, then referrals would constitute a richer hiring pool in precisely the way posited in hypothesis 1. Indeed, if all screening criteria were easily observable to recruiters, the prehire better match theory becomes identical to the richer pool model. As an anonymous reviewer pointed out, a prehire better match account is also a richer pool argument, but one where the referral pool is richer than the nonreferral pool in *unobservable* factors. Thus, choosing between richer pool (hypothesis 1) and the prehire version of better match process comes down to whether the referral applicant pool is richer than the nonreferral pool in observables (hypothesis 2) or unobservables (hypotheses 3–6).

As is always the case in arguments involving unobservables, factors that are posited to be unobservable to screeners are also likely to be unobservable to us as analysts. This raises the issue of our specification of the models in tables 7 and 8. If there are factors that are easily observable to recruiters, but that we do not observe (or measure poorly) in our models, then the dummy variable for referral status in tables 7 and 8 will pick up the effect of the omitted factor. Since we can never be sure that we have included all the relevant factors in our models, we will be at risk of confounding richer pool and prehire better match processes.

Recruiters do appear to prefer referrals even after controlling the ob-

versus not hired) using the Heckman (1979) model. Controlling selection does not introduce significant differences between referrals and nonreferrals in either turnover or wage growth.

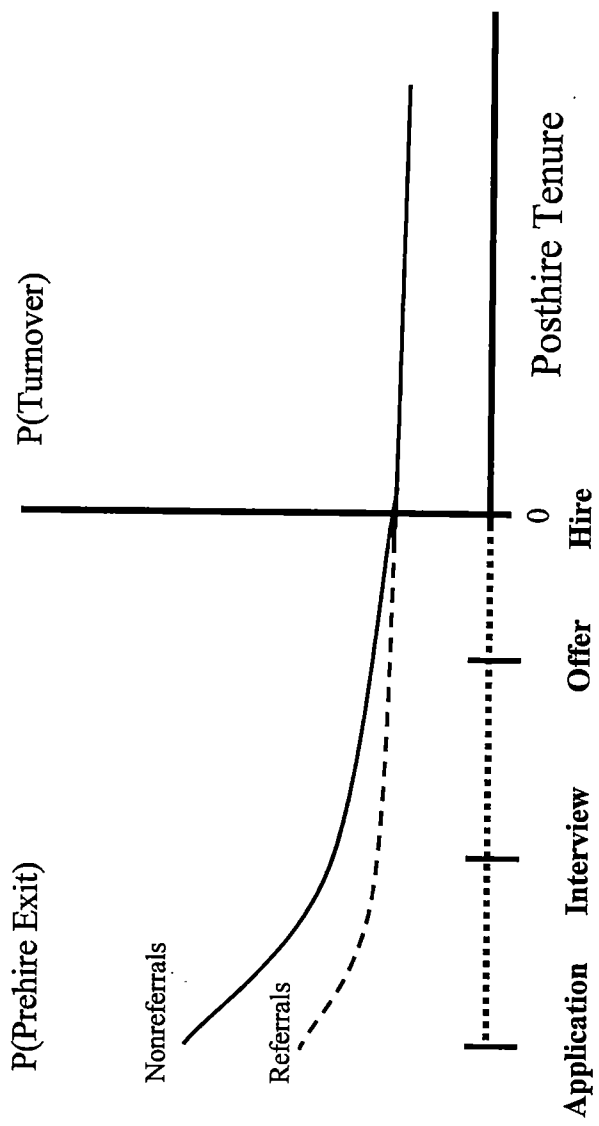


FIG. 4.—Graphical representation of turnover in a prehire better match theory

servable factors in model 2 of tables 7 and 8). This is consistent with a key tenet of a prehire better match account: that referrals are superior to nonreferrals on hard-to-observe characteristics. However, we also think that it is plausible that we have omitted a key variable from these models. Although the evidence is indirect, the data suggest that referrals interview better than nonreferrals (for evidence of this in a retail bank, see Fernandez and Weinberg [1997, pp. 895–98]). At application, referrals are 1.1 times more likely to receive interviews than nonreferrals (64.8% vs. 57.5%). However, referrals' advantages are most manifest post interview. Among interviewees, referrals are over 1.5 times more likely to receive a job offer than are nonreferrals (18.3% vs. 11.6%). Performance during interviews is something recruiters say they place greater emphasis on during screening, but we cannot control for that in these models. Therefore, we think it plausible that the significant coefficients for referral status in both the interview and offer equations reflect a referral advantage in "soft" interviewing skills (on "soft-skills," see Kirschenman and Neckerman [1991]).⁴²

While we cannot easily dismiss the prehire version of the better match story, it is worth noting that our results do place some limits on how such an account could work in this setting. Figure 4 implies that referrals' advantages in terms of observables and unobservables should be greater at earlier rather than later phases of the hiring process. Yet, in our prehire analyses of hypotheses 3–6, we found no evidence that referrals' knowledge advantage over nonreferrals is greater at earlier rather than later stages of the screening process. While we clearly have not exhausted the list of possible unobservable factors, the information differences we did study include ones that would be commonly discussed as "unobservables" in a typical posthire study. To the extent that referrals have better "unobservables," at least in this setting, we can eliminate from consideration the factors listed in table 3. Also, in light of the evidence that PCHR makes no attempt to gather upstream information from the referrer (hypotheses 7–8), we can also eliminate from consideration one of the most important paths through which typically "unobserved" factors are presumed to affect the hiring process.

While we may have narrowed the list of possible omitted unobservable factors in this study, the conservative conclusion to draw is that referral status is proxying things that are unmeasured by us but observable to recruiters (i.e., omitted factors), as well as factors that are unobservable to recruiters. While the question of whether the pool is richer in observables or unobservables might be of theoretical interest, it is important to

⁴² Recall that PCHR often conducts short interviews before deciding whether to pass on candidates to the hiring manager.

TABLE 10

KAPLAN-MEIER SURVIVAL ESTIMATES AT THE MIDWESTERN SITE FOR REFERRERS' STATUS

	Nonreferrals (1)	All Referrals (2)	Referrer Stays (3)	Referrer Leaves (4)
1 month	98.06	97.06	96.79	100.00
3 months	90.97	93.53	93.50	90.00
6 months	81.94	86.47	86.55	80.53
9 months	74.19	74.71	77.14	52.56
1 year	66.39	70.57	73.41	49.06
18 months	55.87	66.09	69.26	43.74
2 years	47.82	53.71	57.99	31.45
<i>N</i> of terminations	87	75	57	18

NOTE.—Log-rank test col. 1 vs. 2: $\chi^2 = 2.30$; $df = 1$; $P < .1292$. Log-rank test col. 1 vs. 3 vs. 4: $\chi^2 = 7.96$; $df = 2$; $P < .0187$. Log-rank test col. 3 vs. 4: $\chi^2 = 7.37$; $df = 1$; $P < .0066$. Log-rank test col. 1 vs. 3: $\chi^2 = 4.37$; $df = 1$; $P < .0422$. Log-rank test col. 1 vs. 4: $\chi^2 = 1.61$; $df = 1$; $P < .2045$.

realize that the mechanism that generates social capital returns to the firm are identical in either case. Whether referrals' advantages are manifest to recruiters (on observables) or operate behind their back (via unobservables), both imply that fewer screens are required to hire referral than nonreferral applicants. Both of these processes will yield the firm returns in the form of savings on recruiting costs. Consequently, we will suppress the distinction between the "mechanism 1 + mechanism 2" prehire better match account and the richer pool model when we discuss the company's return on social capital via savings in screening costs.

Mechanism 5: Posthire Social Processes

The social enrichment account of referral hiring argues that referral ties are a means by which referring employees seek to improve their social environment on the job. In this model, the connection between the new hire and the job is enriched by the existence of a prior friend or acquaintance that might ease the transition to a new job setting. We argued that to the extent that referring relationships have a social component, we should find interdependence between referrers' and referrals' chances of turning over (hypothesis 9).

We tested hypothesis 9 by examining whether the chances of referrals leaving increase when their referrer terminates employment at the company. We first used descriptive methods (examining Kaplan-Meier survival functions) to test whether the survival function for referrals whose referrer leaves is the same as that for referrals whose referrer has stayed with the company (table 10). We then introduced a number of controls

and estimated Cox models predicting the rate of termination (table 11).⁴³ For all these analyses, we coded the "referrer termed" as a time varying variable so that referrals whose referrers have not yet left are coded as 0 until after the referrer leaves (coded as 1). Therefore, the forth column of table 10 reports the Kaplan-Meier estimates of the referrals' survival chances at one month, three months, six months, nine months, and so on after the referrer leaves the firm.

Overall, referrals and nonreferrals do not differ significantly in their propensity to terminate (see the first and second columns of table 10; $P < .1292$; $\chi^2 = 2.30$; $df = 1$). Comparing columns 1 and 3 shows that referrals whose referrers stay are likely to survive longer than nonreferrals ($P < .0422$; $\chi^2 = 4.37$; $df = 1$), although the survival functions of nonreferrals and referrals whose referrer leaves (cols. 1 and 4) are not statistically different from one another ($P < .2045$; $\chi^2 = 1.61$; $df = 1$). However, distinguishing between the third and fourth columns shows statistically reliable evidence of interdependence between referrals' and referrers' turnover. Consistent with hypothesis 9, referrals whose referrer has left show a lower survival rate than referrals whose referrer has stayed ($P < .0066$; $\chi^2 = 7.37$; $df = 1$).

Table 11 presents several Cox models. The first column shows that referrals are not significantly different from nonreferrals in their propensity to turn over. The model in the second column adds a number of individual controls to show that the lack of a referral effect is not due to suppressors. Consistent with hypothesis 9, model 3 shows that referrals whose referrer has termed are more likely to terminate than nonreferrals (the reference category), and that referrals whose referrer has stayed are less likely to leave the company than nonreferrals ($P < .05$). Model 4 tests whether these effects are consistent after introducing several control variables. Supporting hypothesis 9, referrals whose referrer has terminated are significantly more likely to leave the firm than nonreferrals. However, after adding controls, the tendency for referrals whose referrers have stayed with the company to stay longer than nonreferrals is no longer statistically reliable. We reran model 5 with "referral whose referrer stays" as the omitted category in order to test whether referrals whose referrer has left show a lower survival rate than referrals whose referrer has stayed. Similar to the results in table 10, we found that the actions of the referrer significantly distinguish among referrals: referrals whose referrer has left show

⁴³ In preliminary analyses, we estimated other survival-time specifications (e.g., exponential survival time, Weibull, and piecewise-exponential models). The results reported here are consistent across these various specifications.

a significantly higher propensity to turn over than referrals whose referrer has stayed with the firm during our study; hypothesis 9 is supported.

It is clear that referral ties continue beyond the hiring process and have effects later on attachment to the firm. While we have found evidence of social enrichment, any or all of three social mechanisms could operate. (1) Referrers who leave may convey information about external job opportunities back to referrals. (2) The leaving of the referral in itself may be a piece of information prompting the referral to re-evaluate her own satisfaction with the current job. (3) The referrer's exit may reduce the quality of the work environment to an unsatisfactory level. In light of the evidence that a referrer's staying does not reduce turnover relative to nonreferrals (table 11, col. 4), the latter explanation of altering the work environment seems to be a weak candidate. Distinguishing between the first two explanations remains a task for future research.

We looked at the reciprocal effect that referrals had *upstream* on their referrers' turnover. If the social environment is enriched for referrals, it stands to reason that the referrers might also derive a benefit and that the loss of this benefit might prompt turnover. We explored the data for evidence of an upstream effect—referral turnover increasing the chances of referrer turnover (see table 11, cols. 5 and 6). Unfortunately, the study design hampers our ability to mount a robust test of this effect. In order to avoid problems of left censoring, we only looked at referrers who were themselves hired in our two-year hiring window ($N = 82$). Of the 119 referrals these people made, only 18 were hired, and of these, only 7 terminated. Models 5 and 6 add time-varying covariates—one for whether the referred person was hired and one for whether the referral terminated. Although these results are not statistically reliable, the results of these models are suggestive. Despite the lack of statistical power, the effect of the referral terminating in both models is in the predicted direction and approaches acceptable levels of statistical significance ($P < .08$, one-tailed test).

As a final bit of evidence regarding the social enrichment process in this setting, we interviewed PCHR personnel about what they thought were important posthire determinants of retention. PCHR staff are aware of the potential benefits of social enrichment processes. Prior to our study, they had been considering introducing a formal "buddy" system, where long-time employees are paired with new hires as a means of reducing turnover. The underlying theory is identical to that of the social enrichment process. As one trainer put it, "In our culture, it really matters that you have a friend on the job. It really helps in making the job more comfortable." This trainer did not specifically say that she thought the referrer might play an important role in this process. But when we probed about

TABLE 11
COX REGRESSION MODELS PREDICTING TERMINATION AT MIDWEST SITE

	1	2	3	4	5	6
Referral	-.238 (.157)	-.081 (.181)	-.348* (.167)	-.227 (.196)	-.080 (.182)	-.224 (.196)
Referrer terms ^a690* (.281)	.691* (.296)		.684* (.301)
Gender (1 = male)113 (.238)		.091 (.240)	.121 (.240)	.099 (.240)
Marital status (1 = married)		-.360* (.189)		-.400* (.192)	-.360* (.189)	-.398* (.192)
Education (1 = BA)		-.315 (.280)		-.364 (.287)	-.314 (.276)	-.367 (.285)
Computer experience (1 = yes)225 (.246)		.197 (.248)	.209 (.245)	.182 (.248)
Foreign language (1 = yes)503* (.235)		.527* (.235)	.506* (.235)	.531* (.235)
No. of previous jobs037 (.077)		.032 (.077)	.040 (.077)	.035 (.078)

Customer service experience (in months)	-.003 (.002)	-.003 (.002)	-.002 (.002)	-.002 (.002)
Bank experience (in months)007* (.004)	.009* (.004)	.006 (.004)	.008* (.004)
Nonbank experience (in months)	-.008* (.004)	-.007* (.004)	-.008* (.004)	-.007* (.004)
Nonbank experience, squared000027* (.0000008)	.000025* (.0000009)	.000027* (.0000008)	.000025* (.0000008)
Working at application102 (.215)	.056 (.218)	.110 (.219)	.059 (.222)
Last wage	-.030 (.043)	-.024 (.044)	-.029 (.043)	-.023 (.044)
Referred person hired ^a			-.502 (.645)	-.408 (.654)
Referred person terms ^a			1.136 (.744)	1.106 (.739)
χ^2	2.30	26.19	7.60	35.16
<i>df</i>	1	13	2	16
<i>P</i> <1292	.0160	.0224	.0038
<i>N</i>	325	270	318	267
Terminations	163	132	161	130

NOTE.—SEs are given in parentheses.

^a Time-varying covariate.

* *P* < .05, one-tailed test.

the possibility that the referrer might be acting as an informal "buddy," she said that this was "probably right."

THE RETURNS TO SOCIAL CAPITAL

Our analyses to this point have identified the paths through which the firm's investment in the social contacts of its employees operates in the hiring of new employees. We have found evidence of both the richer pool and social enrichment processes, but very little evidence of the posthire better match explanation of referral hiring at work in this firm. We now turn to the question of the dollar returns associated with these processes.⁴⁴

We obtained cost estimates of various recruiting and training activities from PCHR personnel. PCHR social capital investment is \$10 for each referral who is interviewed and \$250 for each referral who is hired and remains with the firm 30 days. This is a net investment, since PCHR does not reduce any of their other recruiting activities due to the fact that the referral program is in place.⁴⁵ Although the number of PCSR job openings may vary over time, PCHR is constantly advertising for and screening PCSR candidates.

PCHR accounts for their screening expenses as follows. Each applicant screen (paper screening plus short telephone interview) costs \$7.00. Each interview (conducted by one person from PCHR plus two line managers) costs \$110. For a referral, the interview cost is \$120, reflecting an additional \$10 paid as a bonus to the referrer. Offering a job, including administrative cost, background check, and drug check, costs \$200. PCHR accounts for advertising costs at the hire stage. Ads cost \$800 per hire, and administrative costs add an additional \$400 per hire. All new hires are required to go through seven weeks of classroom training and two weeks of on-the-job training. The wages and benefits paid to each new hire during training, \$3,930, plus the cost of training—materials and trainers' time—\$1,012, for a total of \$4,942.

The richer pool arguments predict that employers will save on screening costs due to the fact that referrals are more appropriate candidates at application. This implies that fewer screens will be required to produce

⁴⁴ As an anonymous reviewer pointed out, a full analysis of social capital returns would require us to compare the returns generated under the current program with a hypothetical null program that paid zero, something we cannot do here. Our goal in this exercise is more limited. The cost accounting numbers we report below traces the expenses that PCHR experiences under its current practices. While we cannot compare these expenses to other hypothetical systems, we can compare the costs and benefits associated with hiring referrals and nonreferrals under their current arrangement and cost-allocation scheme. For what it is worth, this is the standard that PCHR managers use in judging whether their money was well spent.

⁴⁵ This is important for establishing the baseline cost figures.

TABLE 12

DOLLAR SAVINGS IN SCREENING COSTS ASSOCIATED WITH HIRING VIA THE "RICHER POOL" MECHANISM FOR EACH STAGE OF THE HIRING PROCESS

	Referrals	Nonreferrals	Savings
Application screening:			
Cost per screen	7.00	7.00	...
Screens per candidate interview	1.547	1.744	...
Cost per candidate interview	10.83	12.21	1.38
Interview:			
Cost per interview	120.00	110.00	...
Interviews per candidate offer	5.494	8.645	...
Cost per candidate offer	659.34	950.99	291.65
Offer:			
Cost per offer	200.00	200.00	...
Offers per candidate hire	1.064	1.110	...
Cost per candidate hire	212.87	221.94	9.07

NOTE.—All costs are given in dollar amounts.

one hire among referrals than nonreferrals. In calculating the firms' return on investment, we first decompose the savings into the screening, interview, and offer stages (table 12). Beginning with the screening stage, 64.8% of referrals are granted interviews, compared with 57.5% of nonreferrals, a difference that is statistically significant (see discussion of hypothesis 7 above). This implies that 1.547 (1/648) screens are required to produce one interview for referrals, compared with 1.744 (1/575) screens for nonreferrals. At an average cost of \$7.00, referrals cost \$1.38 less than nonreferrals to produce one interview. This suggests that PCHR is losing money at the application screening stage: the \$1.38 savings due to referrals being from a richer pool of applicants does not outweigh the additional outlay of \$10 to the referrer. In contrast, referrals show a considerable advantage at the interview stage. Among interviewees, referrals are offered jobs at a significantly higher rate than nonreferrals: 18.2% versus 11.6%. In terms of interviews required to produce one offer, among referrals, 5.494 interviews are required versus 8.645 interviews for nonreferrals. It costs \$291.65 less to find a hireable person among referral than nonreferral interviewees. Virtually all offers are accepted, but the statistically insignificant difference between referrals' and nonreferrals' acceptance rate (94.0% vs. 90.1%) leads to referrals being \$9.07 less expensive per hire.

While it is instructive to see the cost savings by stage, these numbers cannot simply be added to yield overall cost savings in hiring because they are not expressed in the same units. In table 13, we recalculated the

TABLE 13

PER HIRE DOLLAR SAVINGS ASSOCIATED WITH HIRING REFERRALS VIA THE "RICHER POOL" MECHANISM FOR EACH STAGE OF THE HIRING PROCESS

	Referrals	Nonreferrals	Savings
Application screening:			
Cost per screen	7.00	7.00	...
Total screens per hire	9.043	16.735	...
Total screening cost per hire	63.33	117.15	53.82
Interview:			
Cost per interview	120.00	110.00	...
Total interviews per hire	5.846	9.596	...
Total interview cost per hire	701.75	1,055.29	355.54
Offer:			
Cost per offer	200.00	200.00	...
Total offers per hire	1.064	1.110	...
Total offer cost per hire	212.87	221.94	9.07
Total costs per hire	977.95	1,394.37	416.43
Referral bonus (investment)	250.00
Total costs	1,227.95	1,394.37	166.43*

NOTE.—All costs are given in dollar amounts.

* Net benefit = 66% return on investment.

screening, interview, and offer costs on a per-hire basis. For referrals, the screening costs per hire are \$63.33, interview costs are \$701.75, and offer costs are \$212.87, for a total of \$977.95 per hire. The corresponding figures for nonreferrals are \$117.15, \$1,055.29, and \$221.94, for a total cost per hire of \$1,394.37. The total difference between referrals and nonreferrals is \$416.43 per hire; 85% of the savings are associated with the interview stage. The \$416.43 difference yields a 66.6% return on the firm's \$250 incremental outlay in the form of the referral bonus. Thus, the firm's social capital investment is justified based on the prehire richer pool process.

We next consider the posthire better match process. As we argued above, to the extent that there are returns associated with better posthire matches, they should manifest themselves in referrals showing lower turnover rates in this setting.⁴⁶ Although posthire turnover outcomes show no significant differences between referrals and nonreferrals, we examined the cost implications of the statistically insignificant difference of 4.2%.⁴⁷

⁴⁶ While we have not examined performance outcomes, PCHR personnel are convinced that turnover costs would dominate any performance differences by recruitment source.

⁴⁷ In all these calculations, we are using the one-year Kaplan-Meier point estimates from table 10.

PCHR calculates its posthire costs (consisting of advertising, administrative, and training costs) to be \$6,142. For every 100 people hired in each category, 29.4 referrals and 33.6 nonreferrals will not survive to one year. Thus, the annual per person costs of replacing these losses is \$1,808 for referrals ($.294 \times \$6,142$) and \$2,064 ($.336 \times \$6,142$). This yields cost savings of \$257 in favor of referrals. So, our best estimate of turnover differences associated with recruitment source shows no practical or statistically reliable return on the \$250 investment vis-à-vis the better match mechanism. In fact, the data for the other phone centers (table 9) show that the return via this process could easily be negative: two of the other phone centers show lower turnover rates for *nonreferrals* at one year.

In contrast, there are much larger financial differences associated with the social enrichment process. Referrals do not differ from nonreferrals in turnover, but there is significant heterogeneity among referrals in turnover depending on the behavior of their referrer. Our estimate of the difference in turnover probability between referrals whose referrer stays versus goes is 24.3% (73.4 vs. 49.1; table 10). This implies that referrals whose referrer leaves have an annual replacement cost of \$3,129, not including the cost of replacing the referrer. Referrals whose referrer stays have an annual replacement cost of \$1,633. For every new hire that can be converted from the "referrer leaves" to the "referrer stays" column of table 10, PCHR can save \$1,496 in replacement costs. However, it is important to note that these are *potential* savings: the bank does not currently realize them. The \$1,496 figure, however, does show that the bank has a large incentive to try to change the behavior of referrers because of the downstream social effects on referrals' turnover propensity. Moreover, this is probably a conservative estimate since downstream effects may extend beyond referral chains of length one. Indeed, we found that 43 referral hires became referrers themselves, successfully producing 5 referral hires in our two-year hiring window.

We conclude this discussion of the returns to social capital by considering the practical implications of our findings. Much as basic knowledge of capital markets allows one to design investment instruments that produce financial returns, we should be able to craft investment strategies for those seeking to trade in social capital. Indeed, the extent of our understanding of these processes is likely to be revealed in our ability to offer such concrete policy recommendations. In light of the large unrealized returns we document above, we focus our suggestions on the social enrichment process.

At the most general level, in attempting to use the social enrichment process for its own ends, PCHR should find ways to increase the referrer-referral relations that cut in its favor. First, the bank could bias their

hiring of referrals in favor of referrers who are likely to stay with the bank, for example, by giving special preference to referrals from long-time employees. This could have two effects, a direct effect of decreasing turnover propensity via the better match mechanism, and an indirect effect of increasing the social support available to the referral. Of course, looking upstream at the characteristics of the referrers will increase the screening cost of hiring referrals. Whether the benefits from reduced turnover via these processes would justify the added costs is an empirical question.

Second, the line managers could be more responsive to the sentiments of referrers, people who are likely to improve the working environment of those they have referred. Our findings suggest that efforts at improving referrers' attachment to the firm will likely pay dividends in the form of reduced downstream turnover. And, to the extent that referrers are originators of referral chains, their opinions are likely to have a disproportionate effect on others.

Third, management could change the timing and structure of the payout to the referrer. One crude change would be lengthening the period the referral must stay before the referrer receives the payout from the current one month. Whether changes along this margin will significantly affect referrers' behavior is an empirical question. Alternatively, the firm's management might seek to fine-tune the payout. Referrers could receive a bonus for every month that the referred person stays with the company. This would have the effect of rewarding those social enrichment actions (e.g., "buddying") that result in higher retention. Another option would exploit the fact that referrals tend to refer others. Modeling the sales practices of some direct sales organizations (see Biggart 1990), PCHR might pay referrers an additional small bonus based on the retention of referrals made by the people they refer. Of course, administering such a system would no doubt add to the cost of hiring from the referral pool and, therefore, will come at the expense of the savings due to the richer pool mechanism.

Finally, managers could try and stem the downstream effects of referrers who leave. Managers could identify and then target retention efforts at referrals that are at high risk of turnover because their referrers have left. In contrast to the strategies (discussed in this article) where the firm harnesses employees' existing social connections for its own ends, this strategy amounts to investing in the selective *breaking* of social ties. While this is certainly a way of managing social capital for the firm's benefit, it must be recognized that such attempts to reshape their employees' social networks run the risk of being perceived by organizational participants as having gone beyond the line of legitimate management activities.

SUMMARY AND CONCLUSION

We argued that a common organizational practice—the hiring of new workers via employee referrals—provides key insights into the notion of social capital. Employers who use such hiring methods are quintessential “social capitalists,” viewing workers’ social connections as resources in which they can invest in order to gain economic returns in the form of better hiring outcomes. We identified three ways through which such returns might be realized: the richer pool, the better match, and the social enrichment mechanisms. We developed a set of falsifiable hypotheses that distinguish among these accounts. Using unique data on hiring from a bank’s credit card phone center, we found support for the richer pool process. However, we have found scant evidence for the posthire better match story, which explains referral hiring as due to improving the firm’s ability to pluck socially isolated individuals from the pool of applicants. We did, however, find evidence supporting the social enrichment process. Consistent with our prediction, we observed interdependence of turnover between referrers and referrals, a process that is not predicted by the socially atomistic better match theory.

We asserted that if social capital is to be more than a metaphor, analysts must identify the investment costs, the rate of return, and the means by which returns are realized. Using unique company data on the dollar costs of screening, hiring, and training, we found that the firm’s investment in the social capital of its employees yields significant economic returns. These returns are realized by savings in screening costs due to referrals being more appropriate for the job at application (i.e., the richer pool mechanism). The firm’s \$250 investment (in the form of a referral bonus) yields a return of \$416 in reduced recruiting costs, a rate of return of 67%. While there is clear evidence of a net benefit to the firm in recruiting referrals via the richer pool process, we found very little evidence consistent with the better match account. Referrals have no better information about the job than do nonreferrals, and recruiters have no better information about referral than nonreferral applicants. Also, contrary to the predictions of the better match theory, nonreferrals are no more likely to turn over than referrals. Consequently, the better match process does not produce significant returns to the firm’s social capital investment.

Nor did we find returns associated with the social enrichment process. Although referrals and nonreferrals do not differ on average in their propensity to turn over, referrals recruited by employees who stay with the firm are more likely to stay with the company than referrals whose referrer leaves the company. To the extent that the firm can manage this interdependence between referrers and referrals, we estimate that the potential returns to the firm are very large. However, in this case, the firm does not

realize these returns. While this firm has clearly missed an opportunity, we suggested ways in which the firm might better manage the interdependence between referrers and their referrals in order to realize returns via the social enrichment process.

This article has made a number of important contributions to the theoretical bases of economic sociology. First, we see this work as an important contribution to the literature on the socially embedded nature of economic processes (Granovetter 1985). We have sought to clarify the various competing explanations of the referral-hiring phenomenon in a fair, dispassionate, and interdisciplinary manner. We suggest that such comparisons are required if we are to make progress clearing away the theoretical ambiguities in discussions of the socially embedded nature of economic life. Much has been written regarding the relationship between the disciplines of sociology and economics (see Baron and Hannan 1994; Hirsch, Michaels, and Friedman 1987). Economists have tended to adopt a "clean model" approach focusing on the logical implications of theory, while sociologists have been guided by a "dirty hands," more empirically grounded approach to theory. As we see it, both approaches have serious limitations unless there is genuine engagement across the disciplinary divide. Especially in the context of a burgeoning field of economic sociology, empirical research without reference to economists' explanations runs a serious risk of preaching only to the choir of the sociological faithful and forgoes the potentially valid insights of the disfavored outgroup. Sociologists can do better when considering economists' theories of labor market processes (see Baron and Hannan 1994). In this article, we have given serious consideration to economists' favored explanation—the better match account—of the referral hiring phenomenon. We offer this article as a "high road" attempt at interdisciplinary engagement with labor economics.

For our economist colleagues, we would like to argue for the benefits of an empirically grounded, case-study approach for shedding light on concrete organizational processes. While the pure theory approach may have the advantage of simplifying phenomena in order to render them more easily understandable, ungrounded theories can blind one to empirically important competing processes. One of the earliest proponents of the better match theory missed clear traces of the social enrichment process in his own data. Ullman (1966) noted that some employers avoid hiring via referrals because of "problems with cliques." The possibility that "problems with cliques" might indicate that posthire social processes between referrers and referrals are not guaranteed to work in the employers' favor did not occur to Ullman. Ullman chose to focus on the turnover implications of referral hiring, "black-boxing" the information transfer process at the heart of the better match process, and to ignore "problems with cliques." The fact is that the discipline of economics has been blind

to the implications of "problems with cliques" for referral hiring. And in the 30 years of research after Ullman published his paper, our research is the first to open up the information transfer processes that have been black boxed by their better match theory and to empirically address competing models of the referral hiring phenomenon. We suggest that it is only by close study of a particular case that we have been able to sharpen the theoretical implications of the various theories to the point that they can be distinguished analytically.

The second major theoretical contribution we offer is with respect to the literature on social capital. In this article, we have sought to take the concept of social capital out of the metaphorical realm. We accomplished this goal by tracing the levels of both investment and returns in real dollar terms and by associating these with concrete social processes occurring during hiring and employment. Further, we drew out the implications for practice by examining a number of ways in which the firm might garner returns via the social enrichment process. As such, this article moves the concept of social capital to an unprecedented level of theoretical clarity and empirical specificity, thus raising the bar for future work employing the idea of social capital.

Finally, this work also has important methodological implications for future research. We have significantly advanced understanding of the important phenomenon of referral hiring by surfacing the role of hitherto neglected features. While we cannot address the generalizability of these findings with our approach, it is important to realize that for many of the hypotheses we test here, *no empirical evidence has ever been offered*. Especially in a crowded field like economic sociology where the competing theories often yield similar predictions, it is only by close study of particular cases that we will be able to sharpen the theoretical implications of the various theories to the point that they can be distinguished analytically. Therefore, we argue that for this area the most productive research strategy is one that has as its goal *depth* of knowledge of particular cases, before pursuing questions of the *breadth* of knowledge, that is, the generalizability of these processes across settings. While we would expect that there will be some contingency in the ways the referral hiring process works in different settings, findings from this research will be very useful in focusing the kinds of information we should pursue in broad-gauge research designed to represent populations of organizations. For example, when studying hiring using organizational surveys, our research suggests that asking recruiters whether they contact referrers would shed light on the better match process. Likewise, for labor market-side surveys, it would be theoretically beneficial to ask referrals whether referrers explained the job to them. For surveys of hired workers, asking referrals what they expect from referrers, and whether anyone (including referrers)

may have helped them in a new job, would be helpful for measuring post-hire social enrichment. Thus, our empirically grounded, case-study approach has also served to illuminate directions for future research. By shedding light on the concrete organizational processes by which firms may harness their employees' social networks, we suggest that this article should stand as a model for the further theoretical development in the field of economic sociology.

REFERENCES

- Attewell, Paul. 1987. "Big Brother and the Sweatshop: Computer Surveillance in the Automated Office." *Sociological Theory* 5:87-99.
- Bailey, Thomas, and Roger Waldinger. 1991. "Primary, Secondary, and Enclave Labor Markets: A Training Systems Approach." *American Sociological Review* 56: 432-45.
- Baron, James N., and Michael T. Hannan. 1994. "The Impact of Economics on Contemporary Sociology." *Journal of Economic Literature* 32:1111-46.
- Biggart, Nicole W. 1990. *Charismatic Capitalism: Direct Selling Organizations in America*. Chicago: University of Chicago Press.
- Blau, Gary J. 1985. "Relationship of Extrinsic, Intrinsic, and Demographic Predictors to Various Types of Withdrawal Behaviors." *Journal of Applied Psychology* 70 (3): 442-50.
- Breaugh, James A., and Rebecca B. Mann. 1984. "Recruiting Source Effects: A Test of Two Alternative Explanations." *Journal of Occupational Psychology* 57:261-67.
- Bridges, William P., and Wayne J. Villemez. 1986. "Informal Hiring and Income in the Labor Market." *American Sociological Review* 51:574-82.
- Burt, Ronald S. 1992. *Structural Holes: The Social Structure of Competition*. Cambridge, Mass.: Harvard University Press.
- . 1998. "The Network Structure of Social Capital." Paper presented at a conference on Social Networks and Social Capital, Duke University, October 13.
- Coleman, James S. 1988. "Social Capital in the Creation of Human Capital." *American Journal of Sociology* 94:S95-S120.
- Corcoran, Mary, Linda Datcher, and Greg J. Duncan. 1980. "Information and Influence Networks in Labor Markets." Pp. 1-37 in *Five Thousand American Families: Patterns of Economic Progress*, edited by Greg J. Duncan and James N. Morgan. Ann Arbor: Institute for Social Research.
- Datcher, Linda. 1983. "The Impact of Informal Networks on Quit Behavior." *Review of Economics and Statistics* 65:491-95.
- De Graaf, Nan Dirk, and Hendrik Derk Flap. 1988. "'With a Little Help from My Friends': Social Resources as an Explanation of Occupational Status and Income in West Germany, the Netherlands, and the United States." *Social Forces* 67:452-72.
- Efron, Bradley, and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap*. New York: Chapman & Hall.
- Fernandez, Roberto M., and Nancy Weinberg. 1997. "Sifting and Sorting: Personal Contacts and Hiring in a Retail Bank." *American Sociological Review* 62:883-902.
- Ferris, G. R., and T. A. Judge. 1991. "Personnel/Human Resources Management: A Political Perspective." *Journal of Management* 17:447-88.
- Granovetter, Mark. 1981. "Toward a Sociological Theory of Income Differences." Pp.

- 11–48 in *Sociological Perspectives on Labor Markets*, edited by Ivar Berg. New York: Academic Press.
- . 1985. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91:481–510.
- . 1995. *Getting a Job: A Study of Contacts and Careers*, 2d ed. Chicago: University of Chicago Press.
- Grieco, Margaret. 1987. *Keeping It in the Family: Social Networks and Employment Chance*. London: Tavistock.
- Halcrow, Allan. 1988. "Employees Are Your Best Recruiters." *Personnel Journal* 67: 42–49.
- Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 45:153–61.
- Hirsch, Paul, Stuart Michaels, and Ray Friedman. 1987. "'Dirty Hands' versus 'Clean Models': Is Sociology in Danger of Being Seduced by Economics?" *Theory and Society* 16:317–36.
- Jovanovic, Boyan. 1979. "Job Matching and the Theory of Turnover." *Journal of Political Economy* 87 (5): 972–90.
- Kalleberg, Arne L., David Knoke, Peter V. Marsden, and Joe L. Spaeth. 1996. *Organizations in America: Analyzing Their Structures and Human Resource Practices*. Thousand Oaks, Calif.: Sage Publications.
- Kirnan, Jean Powell, John A. Farley, and Kurt F. Geisunger. 1989. "The Relationship between Recruiting Source, Applicant Quality, and Hire Performance: An Analysis by Sex, Ethnicity, and Age." *Personnel Psychology* 42:293–308.
- Kirschenman, Joleen, and Kathryn M. Neckerman. 1991. "'We'd Love to Hire Them But . . .': The Meaning of Race for Employers." Pp. 155–74 in *The Urban Underclass*, edited by Christopher Jencks and Paul E. Peterson. Washington, D.C.: Brookings Institution.
- Lane, Julia, and Michael Parkin. 1998. "Turnover in an Accounting Firm." *Journal of Labor Economics* 16:702–17.
- Lepper, Mark R., and David Greene. 1978. *The Hidden Costs of Reward: New Perspectives on the Psychology of Human Motivation*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Lin, Nan, Walter Ensel, and John Vaughn. 1981. "Social Resources and Strength of Ties: Structural Factors in Occupational Status Attainment." *American Sociological Review* 46:393–405.
- LoPresto, Robert. 1986. "Recruitment Sources and Techniques." Pp. 13–1–13–26 in *Handbook of Human Resource Administration*, 2d ed. Edited by Joseph J. Famularo. New York: McGraw Hill.
- Marsden, Peter V. 1994. "The Hiring Process: Recruitment Methods." *American Behavioral Scientist* 37:979–91.
- Marsden, Peter V., and Karen E. Campbell. 1990. "Recruitment and Selection Processes: The Organizational Side of Job Search." Pp. 59–79 in *Social Mobility and Social Structure*, edited by Ronald L. Breiger. New York: Cambridge University Press.
- Miller, Shazia Rafiullah, and James E. Rosenbaum. 1997. "Hiring in a Hobbesian World: Social Infrastructure and Employers' Use of Information." *Work and Occupations* 24:498–523.
- Mincer, Jacob. 1974. *Schooling, Experience, and Earnings*. New York: National Bureau for Economic Research and Columbia University Press.
- Montgomery, James D. 1991. "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis." *American Economic Review* 81:1408–18.
- Mortensen, Dale T. 1988. "Matching: Finding a Partner for Life or Otherwise." *American Journal of Sociology* 94:S215–S240.

- Mortensen, Dale T., and Tara Vishwanath. 1994. "Personal Contacts and Earnings: It Is Who You Know!" *Labour Economics* 1:187-201.
- Myers, Charles A., and George P. Shultz. 1951. *The Dynamics of a Labor Market*. New York: Prentice-Hall.
- Neckerman, Kathryn, and Roberto M. Fernandez. 1998. "Keeping a Job: Network Hiring and Turnover in a Retail Bank." Unpublished manuscript. Columbia University, Department of Sociology.
- Petersen, Trond, Ishak Saporta, and Marc-David Seidel. 2000. "Offer a Job: Meritocracy and Social Networks." *American Journal of Sociology*. In press.
- Pfeffer, Jeffrey. 1989. "A Political Perspective on Careers: Interests, Networks, and Environments." Pp. 380-96 in *Handbook of Career Theory*, edited by M. B. Arthur, D. T. Hall, and B. S. Lawrence. New York: Cambridge University Press.
- Putnam, Robert. 1993. *Making Democracy Work*. Princeton, N.J.: Princeton University Press.
- Rees, Albert. 1966. "Information Networks in Labor Markets." *American Economic Review* 56:559-66.
- Rees, Albert, and George P. Shultz. 1970. *Workers and Wages in an Urban Labor Market*. Chicago: University of Chicago Press.
- Reichers, Arnon E. 1987. "An Interactionist Perspective on Newcomer Socialization Rates." *Academy of Management Review* 12 (2): 278-87.
- Saks, Alan M. 1994. "A Psychological Process Investigation for the Effects of Recruitment Source and Organizational Information on Job Survival." *Journal of Organizational Behavior* 15:225-44.
- Saloner, Garth. 1985. "Old Boy Networks as a Screening Mechanisms." *Journal of Labor Economics* 3:255-67.
- Schwab, D. P. 1982. "Recruiting and Organizational Participation." Pp. 103-27 in *Personnel Management*, edited by K. Rowland and G. Ferris. Boston: Allyn & Bacon.
- Sicilian, Paul. 1995. "Employer Search and Worker-Firm Match Quality." *Quarterly Review of Economics and Finance* 35:515-32.
- Simon, Curtis J., and John T. Warner. 1992. "Matchmaker, Matchmaker: The Effect of Old Boy Networks on Job Match Quality, Earnings, and Tenure." *Journal of Labor Economics* 10 (3): 306-29.
- Sørensen, Aage B., and Arne L. Kalleberg. 1981. "An Outline of a Theory of the Matching of Persons to Jobs." Pp. 49-74 in *Sociological Perspectives on Labor Markets*, edited by Ivar Berg. New York: Academic Press.
- Spence, A. Michael. 1974. *Market Signaling: Informational Transfer in Hiring and Related Screening Processes*. Cambridge, Mass.: Harvard University Press.
- Stoops, Rick. 1981. "Employee Referral Programs: Part I." *Personnel Journal* 60:98.
- . 1983. "Employee Referral Programs." *Personnel Journal* 62:690-92.
- Sutton, Robert I., and Meryl Reis Louis. 1987. "How Selecting and Socializing Newcomers Influences Insiders." *Human Resources Management* 26:347-61.
- Swaroff, Phillip, Elizabeth Barclay, and Alan Bass. 1985. "Recruiting Sources: Another Look." *Journal of Applied Psychology* 70 (4): 720-28.
- Taylor, M. Susan, and Donald W. Schmidt. 1983. "A Process Oriented Investigation of Recruitment Source Effectiveness." *Personnel Psychology* 36:343-54.
- Tuma, Nancy. 1976. "Rewards, Resources, and the Rate of Mobility: A Nonstationary Multivariate Stochastic Model." *American Sociological Review* 41:338-60.
- Ullman, Joseph C. 1966. "Employee Referrals: Prime Tool for Recruiting Workers." *Personnel* 43:30-35.
- Wanous, J. P. 1980. *Organizational Entry: Recruitment, Selection, and Socialization of Newcomers*. Reading, Mass.: Addison-Wesley.
- Winship, Christopher, and Robert D. Mare. 1992. "Models for Sample Selection Bias." *Annual Review of Sociology* 18:327-50.

The Power to Punish: Discretion and Sentencing Reform in the War on Drugs¹

Rodney L. Engen
North Carolina State University

Sara Steen
Vanderbilt University

This article addresses whether and how organizational processes and the exercise of discretion affect the relationship between sentencing reforms and sentencing outcomes. Organizational perspectives are drawn upon to develop competing hypotheses regarding courts' adaptations to and implementation of reforms under sentencing guidelines. To test these hypotheses, changes in the charging and sentencing of drug offenders in Washington State from 1986 to 1995 are examined following a series of sentencing reforms. The findings suggest that changes to the sentencing laws have resulted in adaptations by courtroom workgroups that are consistent with an *organizational efficiency* model of criminal justice.

INTRODUCTION

Criminal sentencing in the United States has undergone profound changes over the last 20 years, culminating in the enactment of determinate and presumptive sentencing guidelines in many states and at the federal level (Tonry 1996). By restricting judicial discretion and requiring that sentences be based primarily on characteristics of the offense and the offend-

¹ This research was supported in part by a stipend from the North Carolina State University College of Humanities and Social Science. Data and supporting documentation were provided by the Washington State Sentencing Guidelines Commission. The authors are solely responsible for the analyses and interpretation of the data presented here. The authors wish to acknowledge the helpful comments of the anonymous *AJS* reviewers, George S. Bridges, Patricia L. McCall, and William R. Smith for their comments on earlier drafts, and Joachim J. Savelsberg for his thoughtful remarks on an earlier version presented at the 1999 ASA meetings. Direct correspondence to Rodney L. Engen, Department of Sociology and Anthropology, Box 8107, North Carolina State University, Raleigh, North Carolina 27695. E-mail: rlingen@server.sasw.ncsu.edu

er's prior record, guidelines seek to increase uniformity in sentencing, eliminate unwarranted disparities, and, to a large degree, institutionalize principles of just deserts and deterrence as the appropriate basis for criminal sentencing (Savelsberg 1992).

Critics, however, have expressed doubts that sentencing guidelines would achieve these goals (Walker 1993). Miethe (1987), for instance, states that "the national movement toward determinate sentencing is considered the primary example of a reform effort that will have its intended goals circumvented by the *hydraulic displacement of discretion*" (p. 156; emphasis added). Because charging decisions determine sentencing options, discretion and control over sentencing outcomes simply shift from the sentencing stage to the charging stage, which remains formally unregulated (Alschuler 1978; Nagel and Schulhofer 1992; Boerner 1995; Tonry 1996). Furthermore, laws are implemented by communities of actors and in specific organizational contexts (Miethe 1987; Savelsberg 1992; Dixon 1995; Ulmer 1997). If courtroom workgroups (in particular, prosecutors) use those guidelines in the service of their own personal, cultural, political, or organizational goals, the formal goals of sentencing guidelines may be compromised or even contradicted.

While the notion of hydraulic displacement continues to influence contemporary research (e.g., Albonetti 1997; Ulmer 1997), little research has even *attempted* to test its central hypothesis, much less to examine its effects on uniformity in sentencing outcomes or the organizational forces that may drive it (Miethe 1987; Savelsberg 1992). Research addressing these issues would have obvious implications for policy that attempts to regulate the process of punishment. Perhaps more importantly, however, such research has the potential to provide critical insights into the roles of organizational processes and social inequality in formal social control.

With these broad issues in mind, this study addresses the questions of whether and how the exercise of discretion by courtroom workgroups affects the relationship between sentencing reforms and sentencing outcomes. Specifically, does the displacement of discretion under sentencing guidelines undermine the goals of sentencing reforms, as critics contend? If so, how, and toward what ends, do courtroom workgroups exercise their discretion? To examine these questions, we draw upon organizational perspectives and research on courtroom workgroups, identify various goals that may shape the sentencing process, and develop competing hypotheses regarding courts' likely adaptations to reforms under sentencing guidelines. To test these hypotheses, we examine changes in the conviction and sentencing of drug offenders in Washington State, from 1986 to 1995, following a series of changes to the state's sentencing guidelines.

FINDINGS AND LIMITATIONS OF RESEARCH UNDER SENTENCING GUIDELINES

Because one of the principal goals in the creation of sentencing guidelines was to increase uniformity in sentencing, most studies of the impact of sentencing guidelines have compared the effects of legally relevant factors and extralegal offender characteristics on sentencing outcomes. The bulk of this research suggests that guidelines have largely succeeded in achieving judicial compliance, reducing sentencing disparities by sex or race, and increasing uniformity (Moore and Miethe 1986; Nagel and Schulhofer 1992; Frase 1993; Kramer and Steffensmeier 1993; Tonry 1996). However, a number of studies in states with guidelines find that, while legally relevant factors are the major determinants of sentencing outcomes, extralegal factors including race, gender, and mode of disposition (guilty plea versus trial) continue to affect the severity of sanctions (Miethe and Moore 1985; Moore and Miethe 1986; Albonetti 1992, 1997; Kramer and Steffensmeier 1993; Steffensmeier, Kramer, and Streifel 1993; Dixon 1995; Kramer and Ulmer 1996; Ulmer 1997). Similarly, some studies find that social and organizational context (including factors such as bureaucratization, court size, local resources, and political discourse) can impact the use of sentencing guidelines and may condition the effects of other extralegal characteristics (Dixon 1995; Ulmer and Kramer 1996; Albonetti 1997; Ulmer 1997). Importantly, Ulmer (1997) concludes that "the nature and character of justice and formal social control . . . depend as much or more on the processual orders of local courts as they do on the policies and laws of larger-scale state actors" (Ulmer 1997, p. 189).

Most sentencing studies, however, cannot answer the questions that are the focus of the present research: *whether* and *how* organizational processes affect the relationship between sentencing reforms and sentencing outcomes. First, studies that focus exclusively on sentencing outcomes do not address the fundamental criticism of sentencing guidelines, the displacement thesis, which is that *discretion shifts to earlier decision-making stages*, principally charging and plea bargaining. Analyses that "control" for offense characteristics at the point of conviction necessarily mask that hydraulic effect. Second, because most studies are cross-sectional, they do not address the central prediction of the displacement thesis, which is that *prosecutorial behavior will change in response to changes in sentencing laws* (Miethe 1987). Thus, most studies overlook the most important ways in which organizations may adapt to and potentially undermine sentencing reforms, which is by changing behavior at the prosecutorial stage.

To our knowledge, only three scholarly studies, each of which analyzed sentencing in Minnesota, have attempted to do this. Using extensive data

on offenders sentenced before and after guidelines were established in Minnesota, Miethe and Moore (1985) and Miethe (1987) explicitly examined presentencing decisions related to plea bargaining (charge and sentence reductions) and sentencing outcomes, controlling for legally prescribed characteristics (offense characteristics and prior offenses) and legally proscribed factors (social status characteristics and method of conviction). Contrary to the displacement hypothesis, however, they found little evidence that changes in prosecutorial practices, postguidelines, undermined the intent of guidelines. Changes that did occur were, they conclude, due mainly to changes in case attributes over time. Furthermore, the effects of offender characteristics, and their relationship to charging practices, remained relatively stable or decreased during the first two years under sentencing guidelines. Miethe (1987) concluded that the findings of those studies question the assumption that preguidelines sentencing disparities will reappear as postguidelines charging disparities and suggests that social control mechanisms within courtroom workgroups may limit the exercise of prosecutors' discretion and thereby maintain uniformity.

Contrary to that conclusion, however, Knapp (1987) provides evidence that courtroom workgroups *did* in fact circumvent specific provisions of the Minnesota sentencing guidelines (see also Savelsberg 1992; Frase 1993). That study found that prosecutors in Minnesota manipulated charges in ways that mitigated the impact of sentencing guidelines designed to decrease the use of prison sentences for property offenders. Specifically, they required offenders to plead guilty to multiple counts, thus increasing criminal history scores and increasing the presumptive sentence. Similarly, judges there succeeded in using dispositional departures to avoid sharply increased sentences for many sex offenders, for whom judges generally advocated rehabilitative dispositions. Importantly, as Savelsberg observes, in each example criminal justice actors' substantive concerns regarding appropriate punishment were manifested in organizational adaptations that largely muted the effects of sentencing guidelines.

In sum, extant research offers evidence of the importance of organizational and other extralegal factors in the sentencing process in the context of sentencing guidelines, but little research addresses the hydraulic displacement thesis. Research examining the impact of the movement from indeterminate sentencing to determinate sentencing guidelines on case processing (including stages other than sentencing) is limited to three studies in one state, and to a relatively short period of time after guidelines were introduced. Furthermore, virtually no research has examined the impact of *subsequent* reforms in states that have adopted sentencing guidelines. This study extends the literature on sentencing guidelines and reform and overcomes these limitations by examining both charging (i.e.,

conviction) and sentencing outcomes and by examining adaptations to a series of sentencing reforms over a 10-year period in Washington State.

ORGANIZATIONAL PERSPECTIVES ON CRIMINAL SENTENCING

Despite the theoretical and practical importance of the hydraulic displacement thesis, relatively little theory has been advanced that can provide specific hypotheses about the likely impact of changes to sentencing laws. Toward that end, in this section we examine organizational perspectives on the sentencing process. We use these perspectives to identify different operational goals that may shape that process within court communities and that suggest different responses by court communities to the imposition of external rules such as sentencing guidelines. From these perspectives, we then develop additional hypotheses about how courts are likely to respond to *changes* in sentencing guidelines.

Savelsberg's (1992) analysis of sentencing guidelines provides a useful starting point and conceptual framework. Applying Weber's (1968) distinction between *formal* and *substantive rationality* of law, Savelsberg argues that sentencing guidelines attempt to reverse the substantive rationalization of criminal law and to return to law based on a neoclassical formal rationality.² In his classification of legal systems, Weber (1968) describes formally rational law as law guided by norms intrinsic to the legal system, bound by objectively verifiable and legally relevant facts, and comprising a "gapless" set of rules from which decisions in concrete cases can be deduced. Substantively rational law, in contrast, takes into account "economic, sociological, and ethical criteria" and represents "a desire for 'substantive expedience,' a concern with impact, effect, and ends served by legal action, and adaptation 'to the concrete case'" (Weber [1968] quoted in Savelsberg [1992, pp. 1346–48]). Sentencing guidelines represent formally rational law to the extent that they require decisions to be based exclusively on legally relevant facts, and without regard to substantive concerns about the offender, the perceived causes of criminality, or the anticipated consequences of punishment for the individual offender.

Savelsberg (1992) predicts that the formal rationality implicit in sentencing guidelines will fail within the current legal context. While his argument is multifaceted, two central themes are directly relevant to the

² More precisely, Savelsberg (1992) argues that sentencing guidelines attempt to achieve *sociologically formal rationality of justice*, based on neoclassical principles of deterrence and just deserts, through the mechanism of *logically formal rationality of law*. Savelsberg here adopts Ewing's (1987) distinction between logically formal rationality of legal thought and sociologically formal rationality of justice in Weber's sociology of law.

present study. First, he argues that the limited discretion that judges retain under guidelines, combined with the substantial control over sentences that can be exercised through charging decisions (the hydraulic displacement thesis), provide mechanisms through which formal rationality can be subverted. Second, he argues that criminal justice in the United States is firmly entrenched in substantive concerns. Guidelines "have to be implemented through networks of decision makers in complex administrative environments and by actors with *internalized substantive rationales*" (p. 1361; emphasis added). In arriving at appropriate dispositions, criminal justice organizations and the actors within those organizations will resist restraints on their autonomy and will continue to take into account concerns related to social inequality, social justice, perceived causes of criminal behavior, and the anticipated consequences of sentencing (Savelsberg 1992, pp. 1352–53).³ Thus, charging and sentencing decisions will remain substantively rational under guidelines.

Examples of theoretical and empirical work emphasizing courts' substantive rationality are common in the literature on race and gender disparities in punishment (e.g., Peterson and Hagan 1984; Daly 1987, 1994; Bridges and Crutchfield 1988; Albonetti 1991; Dixon 1995; Steffensmeier, Ulmer, and Kramer 1998). While that research identifies a wide range of concerns as driving disparities in sentencing, the arguments are consistent with the notion that criminal justice organizations are oriented toward achieving particular substantive outcomes (e.g., protecting the public or controlling "problem" populations). Substantive rationality is also prominent in the literature on courtroom workgroups and local legal culture, a central feature of which is the establishment of *going rates*, informal norms that determine decisions in most cases (Sudnow 1965; Rosett and Cressey 1976; Eisenstein and Jacob 1977; Nardulli, Flemming, and Eisenstein 1985). More recently, Ulmer (1997) extends that research tradition to the context of sentencing guidelines, arguing that "court communities are thus policy arenas . . . where two sets of sentencing standards meet—the formally rational ones articulated by guidelines (such as offense severity, prior criminal record) versus the substantive, extra-legal criteria deemed relevant by local court actors and local legal cultures" (p. 4).

In contrast to this emphasis on substantive rationality, some researchers argue that the sentencing process is heavily influenced by courts' need for *organizational efficiency* and that those demands may supersede both formal and substantively rational goals (Packer 1968; Balbus 1973; Hagan

³ An anonymous reviewer noted that implementers' opposition to sentencing guidelines may take one of two forms: (a) resistance to the restraint on their autonomy that logically formal rationality entails; and (b) a commitment to substantively rational concerns that are at odds with the sociological rationality of guidelines.

1989; Dixon 1995). Dixon (1995) traces this perspective on sentencing to "Michels' law": "According to Michels (1915), the imperatives of organizational maintenance deflect the system from the pursuit of formal rational goals and result in the development of operating goals by organizational elites" (p. 1162). Extending this principle to criminal justice organizations, Dixon argues, produces a sentencing theory that "defines efficient case disposition as the operational goal that maintains a stable and orderly sentencing system" (p. 1162). The goal of criminal justice processing, according to this perspective, is to induce guilty pleas as an organizationally efficient means to an end. Thus offenders who plead guilty are rewarded with less severe sentences than they could otherwise have received.

To summarize, we identify three alternative perspectives, or models, of the sentencing process that assume different operational goals of courtroom communities and that suggest different ways in which sentencing reforms may affect sentencing and case processing. First, the *formal legal* model assumes the operational goal of formally rational criminal justice, which is simply to impose sentences that are commensurate with the offenses committed and in accordance with sentencing laws. Second, the *substantive rationality* model assumes organizational commitments to established norms, internalized rationales about appropriate sentences and sentencing criteria, and an operational goal of producing sentences consistent with those substantive concerns. Third, the *organizational efficiency* model assumes the imperative of disposing of cases efficiently and the operational goal of inducing large numbers of offenders to plead guilty.

While probably no theorist would advance any one of these models to the exclusion of the others, it is useful for analytic purposes to treat them as distinct, ideal types, and to explore the implications of each for organizational adaptations to changes in sentencing laws.⁴ The critical distinction between the substantive rationality and organizational efficiency models is that the former is oriented toward the ends of the process (sentences), while the latter is oriented toward the means by which those ends are achieved. Both contradict a formal legal rationality, which is oriented toward the law. As such, these models imply different responses to sentencing reforms. By developing those implications, and examining courts' responses to changes in the law, we hope to shed light on the question of whether criminal justice is organized primarily toward achieving and maintaining formally rational justice, substantively rational outcomes, or efficient case processing.

⁴ It is important to note that these different goals are not necessarily incompatible in practice, and in certain instances they may be complimentary. For example, substantively rational going rates and formally rational sentencing guidelines can both facilitate case processing by simplifying decision making (Eisenstein et al. 1988; Ulmer 1997).

SENTENCING REFORM IN WASHINGTON STATE

In 1984, Washington State implemented its Sentencing Reform Act (SRA), which established specific rules for the sentencing of persons convicted of felony crimes in Washington State. The SRA provides a determinate sentencing model, with relatively narrow presumptive ranges determined by the *seriousness level* (1–15) of the offender's most serious current offense and the *offender score* (0–9+), which represents the offender's prior felony convictions and other concurrent charges. Alternatives to sentences within the presumptive range are limited as specified in the SRA.⁵

Since implementing the SRA, Washington State has amended the sentencing options and the applicable terms of incarceration substantially for offenders convicted of delivering heroin or cocaine.⁶ The first amendment to the SRA that we examine, effective in fiscal year 1988, eliminated the First-Time Offender Waiver (FTOW) for offenders convicted of *delivery* of heroin or cocaine. With the FTOW, offenders who were first-time, non-violent felons had been eligible for a reduced sentence of up to 90 days confinement, plus two years of community supervision, with additional sentence conditions optional. The principal effect of eliminating the FTOW was that, without resorting to an *exceptional* sentence, every convicted dealer would receive a prison sentence within the standard range (12–14 months for a first offender).⁷

A second set of reforms, effective in fiscal year 1990, substantially increased the presumptive sentencing ranges for these offenders by: increasing the seriousness level of delivery of heroin or cocaine from level 6 to level 8, changing the formula for scoring prior or concurrent drug deliveries in the offender score, increasing the mandatory "deadly weapon enhancement" from 18 to 24 months, and introducing a mandatory 24-month sentence enhancement for offenders convicted of narcotics delivery within a "protected zone."⁸ The resulting increases in presumptive sentences are

⁵ "Presumptive" sentencing guidelines, like those in Washington State, require that judges sentence within the specified (i.e., presumptive) range unless specific conditions are met. Sentencing is "determinate" in that judges order a precise term of confinement (in months), from within the presumptive or "standard" range.

⁶ The offense that is the focus of these sentencing reforms, and of our analysis, is "manufacture, deliver or possess with intent to deliver a schedule I or schedule II narcotic," specifically heroin or cocaine (RCW 69.50.01[a][1][i]). We will refer to this as simply "delivery."

⁷ The SRA allows judges to order an "exceptional" sentence, outside the standard range, under "substantial and compelling" circumstances, provided that they justify these exceptions in writing.

⁸ "Protected zones" are schools, school buses, within 1,000 feet of a school ground or school bus route stop, public parks, public transit vehicles, and public transit stop shelters.

dramatic. Applying this to an individual case, an offender with no prior felony convictions who was convicted of one count of delivery of heroin or cocaine would see his sentence increase from a standard range of 12–14 months to a standard range of 21–27 months—essentially doubling the sentence length. As a result of the change in scoring rules, along with the increase in the seriousness level, the same offender convicted of *two* counts of delivery of heroin or cocaine would face a prison sentence of between 36 and 48 months, compared to a presumptive range of between 21 and 27 months prior to 1990.

The third change in sentencing for drug offenders was the result of a 1992 state Court of Appeals case (*State v. Mendoza*). Prior to *Mendoza*, most anticipatory offenses, including drug deliveries, were sentenced at 75% of the presumptive range for a completed offense. In *Mendoza*, the court ruled that conspiracy to deliver narcotics (and by implication, attempted delivery) is an unranked offense (i.e., one that has no seriousness level defined in the SRA). As an unranked offense, *conspiracy to deliver* carries with it, by default, a presumptive sentencing range of 0–12 months, thereby reintroducing the possibility of a nonprison sentence, or even a sentence of *no* additional time in incarceration, for offenders charged with this offense. Furthermore, this rule applies *regardless* of the number of counts for which the offender is convicted, provided the most serious of these is conspiracy to deliver. As a result, an offender—even a repeat offender—convicted of one or more counts of conspiracy to deliver heroin or cocaine could receive a shorter sentence than a first-time offender convicted of a single count of simple possession.

IMPLICATIONS OF REFORMS FOR CASE PROCESSING

Under Washington State's sentencing guidelines, sentences are largely determined, both in kind and severity, by the specific charge for which an offender is convicted and by the offender's criminal history. In this legal context, the discretion afforded prosecutors (more precisely, their control over sentencing outcomes) is enhanced considerably, while the discretion afforded judges is simultaneously reduced. By determining the charges that will be filed and, in most cases, pled to, the prosecutor determines the extent and type of discretion available to the sentencing judge. Charge manipulation in this context therefore makes the sentence *highly* predictable. By filing, or not filing, certain charges, whether unilaterally or as part of a negotiated plea agreement, the prosecutor can be reasonably certain that a particular sentence will result.

As sentencing reforms are enacted that change the guidelines, the eligibility of certain offenders for certain types of sentences changes. As a result, the options available to the prosecutor seeking to affect a particular

outcome change as well. In one sense, though, the options available to prosecutors for affecting sentences are constant. Prosecutors have the discretion to: charge all legally sufficient counts, or some lesser number of offenses; charge the most serious offense possible, or reduce the primary charge to some lesser but necessarily included offense; and make sentencing recommendations to the court. What changes, as a consequence of reforms, is the impact that those discretionary decisions have on the resulting sentence. By first determining the options that are available at different times, either as incentives to defendants to plead guilty or as mechanisms by which to maintain substantively rational punishment, and then examining which of those options are exercised, we can shed light on the questions of whether, and toward what end, discretionary power is used by courtroom workgroups when laws change.

With the elimination of the FTOW in 1988, there were essentially no alternatives to a prison sentence for offenders convicted of delivering narcotics. Prosecutors could recommend a sentence at the bottom of the presumptive range, but given the narrow ranges in place to begin with (12–14 months for a first-time offender), that provided little incentive for the defendant to plead guilty and was no substitute for the FTOW. The only other alternative was to reduce the severity of charges. Reducing the number of counts against the defendant might suffice to induce a guilty plea, but it would still mean a prison sentence and is only an option if there is evidence of multiple offenses. Reducing the primary charge from delivery to simple possession (a level-2 felony), on the other hand, would produce a substantial sentence reduction (0–90 days for a first-time offender, the same sentence range that was possible with the FTOW). Another option would be to reduce the primary offense to an anticipatory offense (attempt, conspiracy, or solicitation to deliver heroin or cocaine), offenses which, up until 1992, were sentenced at 75% of the presumptive range for a completed offense (i.e., 9–10.5 months for a first offender). Depending on how narrowly one interprets the definitions of attempt, solicitation, and conspiracy (hereafter referred to collectively as “conspiracy”), these crimes are implicit in most drug-delivery offenses.

The second set of reforms, effective in 1990, substantially increased the sentence offenders convicted of delivery would necessarily receive, even if judges ordered the minimum sentence possible. What options, then, did prosecutors have as means by which to either maintain current rates of punishment or encourage guilty pleas? Essentially, the same options are available—reducing the number or severity of charges—with the additional option of agreeing to not allege facts that would trigger an additional 24-month enhancement. The relative value resulting from these options, in terms of incarceration time, has changed, however. Whereas, prior to 1990, conviction on a second count of delivering narcotics would

add up to 13 months to the presumptive range, beginning in 1990, that second count would add up to 21 months to the presumptive range. Further, because the presumptive range for delivery roughly doubled, the value of a charge reduction to an anticipatory delivery is also greater than it had been previously. Thus, beginning in 1990, a substantial sentencing reduction could be achieved either by reducing the number of charges or by reducing the severity of the offense from delivery to an anticipatory offense or to simple possession.

Finally, the reintroduction in 1993 of a nonprison sentence, per *Mendoza* (1992), for a delivery-related offense provided prosecutors with yet another option for offenders who could be charged with delivery. By reducing a delivery charge to conspiracy to deliver, prosecutors could reduce a lengthy presumptive prison sentence to a nonprison sentence. That ability gave prosecutors a powerful tool that could then be used either as a “bargaining chip” with offenders or as a mechanism for achieving and maintaining a substantively appropriate level of punishment.

HYPOTHESES

From the theoretical perspectives identified above, we derive competing sets of hypotheses regarding the most likely consequences of these sentencing reforms. In deriving these hypotheses, we again consider the particular operational goals of criminal justice organizations implied by each perspective and the specific options available to prosecutors and courtroom workgroups for achieving those goals subsequent to each change in the law. We state these hypotheses both as general predictions about the implementation of sentencing reforms and as specific predictions about the impact on (1) sentencing outcomes and (2) charging outcomes, and the relation of each to the mode of conviction. While most quantitative research on the impact of sentencing guidelines has focused on the effects of case and offender characteristics on individual-level outcomes, the organizational processes underlying our hypotheses will be more readily observed in the collective outcomes of those processes. Therefore, we state our hypotheses in terms of changes in aggregate-level outcomes. The hydraulic displacement thesis—that courtroom workgroups will manipulate charges in order to circumvent the intent of reforms—is implicit in hypotheses based on substantive rationality and organizational efficiency, so is not developed here explicitly.

General Predictions

The formal legal perspective is useful in developing what are essentially null hypotheses about the role of organizational processes in the imple-

mentation of sentencing reforms. This perspective assumes the operational goal of punishing offenders in accordance with the law and their actual crimes and predicts the pure application of sentencing guidelines and reforms. Reforms based on offense characteristics should therefore result in sentences in accordance with the guidelines. Changes may occur in the severity of charges over time, but only as a result of changes in criminal behavior, and should therefore be unrelated to changes in the law. Organizational processes should not mitigate the effects of sentencing reforms on sentencing outcomes.

If court communities are oriented toward achieving substantively rational outcomes, we would expect charging practices to change in predictable ways and for those adaptations to mitigate the impact of sentencing reforms. When sentencing reforms depart significantly from existing practices, courtroom workgroups will continue to rely on their own informal norms and substantive criteria for appropriate punishment and will modify charging practices so that existing levels of punishment are maintained, or at a minimum, the effects of reforms are muted. Sentencing outcomes, overall, will therefore not be significantly altered by changes in sentencing guidelines.⁹

Finally, the principal focus and operational goal of criminal justice processing may be maintaining organizational efficiency. The mode of conviction—guilty plea versus trial—lies at the center of predictions based on that goal. Because guilty pleas enable the organization to operate efficiently, offenders who plead guilty will be rewarded with charge reductions, resulting in shorter sentences. As the options available for rewarding guilty pleas change, charging practices will change in ways that continue to reward those who plead guilty but that are irrespective of substantive concerns about appropriate punishment for those offenders. That is, the overall severity of charges and sentences will differ by mode of conviction (as offenders who plead guilty are rewarded), even though sentences may increase in severity for *all* offenders. Furthermore, because there is no organizational incentive to change charging practices unless defendants plead guilty, so too will aggregate changes in charging depend upon the mode of conviction. The general assumptions and predictions above lead to the following hypotheses regarding aggregate sentencing outcomes and patterns of convictions over time.

⁹ We do not equate substantive rationality with a motivation to maintain going rates of punishment. However, *stable aggregate rates* of punishment are a logical consequence if substantive criteria are applied consistently over time. That is true regardless of whether stability is the objective and even if sentencing criteria are applied in an inconsistent or discriminatory way at the individual level. However, if the substantive criteria or objectives change over time, aggregate sentencing outcomes would also change. In that case though, we expect those changes to be independent of changes in law.

Hypotheses Regarding Sentencing Outcomes

Formal Legal Rationality

HYPOTHESIS 1.—*The severity of sentences will change in accordance with changes mandated by sentencing reforms.*

HYPOTHESIS 2.—*The severity of sentences will be independent of the mode of conviction.*

Substantive Rationality

HYPOTHESIS 3.—*The severity of sentences will remain relatively stable over time and independent of changes mandated by sentencing reforms.*

HYPOTHESIS 4.—*The severity of sentences will be largely independent of the mode of conviction.*

Organizational Efficiency

HYPOTHESIS 5.—*Offenders convicted at trial will receive more severe sentences than offenders who plead guilty, regardless of changes mandated by sentencing reforms.*

Hypotheses Regarding Charging Outcomes

Formal Legal Rationality

HYPOTHESIS 6.—*Changes in the severity of charges over time will be independent of changes in the law.*

HYPOTHESIS 7.—*Changes in the severity of charges over time will be independent of the mode of conviction.*

Substantive Rationality

HYPOTHESIS 8.—*The severity of charges will change over time in response to sentencing reforms, decreasing as the severity of sentencing for delivery is increased.*

HYPOTHESIS 9.—*Changes in the severity of charges over time will be largely independent of the mode of conviction.*

Organizational Efficiency

HYPOTHESIS 10.—*Offenders who plead guilty will be convicted of less severe charges than will offenders convicted at trial, regardless of changes in sentencing laws.*

HYPOTHESIS 11.—*Among offenders who plead guilty, the severity of charges will change over time as sentencing reforms change the prosecutorial options for inducing and rewarding guilty pleas.*

HYPOTHESIS 12.—*Among offenders convicted at trial, changes in the severity of charges will be independent of changes in the sentencing law.*

METHODS

Data and Measures

This analysis examines conviction and sentencing data for adult offenders sentenced in Washington State between July 1, 1985, and June 30, 1995 (fiscal years 1986 to 1995), for crimes involving possession or delivery of schedule 1 or 2 narcotics, which are primarily heroin and cocaine.¹⁰ These data are collected and maintained by the research staff of the Washington State Sentencing Guidelines Commission (SGC) and are recorded directly from the individual *judgement and sentence* forms submitted monthly by each Superior Court in the state.¹¹

Conviction data are inherently limited for the study of charging decisions. The primary limitation is, of course, that the data do not include the specific type or number of offenses for which offenders were arrested and referred to county prosecutors. As a result, it is impossible to measure charging decisions in any individual case, and thus we cannot observe prosecutors' exercise of discretion at that level. Also, like most studies, we have no data on acquittals or dismissals of charges initially filed by prosecutors, or of prosecutors' declining to file charges in the first place.

Despite these limitations, conviction data are useful as indicators of aggregate-level charging patterns over time. Aggregate characteristics of conviction offenses (e.g. the prevalence of delivery convictions relative to lesser offenses) at any point in time are primarily a product of two things: (1) the kinds of cases referred for prosecution and (2) prosecutors' charging decisions. Changes in conviction offenses over time therefore reflect

¹⁰ While it would be preferable to examine only sentences for heroin or cocaine offenses, this is not possible. Possession of heroin and cocaine are not differentiated in the sentencing law from possession of other schedule 1 or 2 narcotics, so it is impossible to determine the specific type of drug from the conviction offense in those cases. Prior to 1990, the same was true for delivery. However, it is widely held among criminal justice officials in the state that nearly all of these offenses do in fact involve heroin or cocaine. The 1990–95 convictions for delivery of narcotics support that belief; out of 8,313 convictions, 279 (3%) are for other schedule 1 or 2 narcotics. It seems reasonable to assume that the possession convictions, and delivery convictions prior to 1990, include similar proportions by type of drug.

¹¹ Washington Superior Courts are required by the legislature to submit a judgement and sentence form to the SGC for each felony sentence. Given that the data are used for monitoring sentencing and advising the legislature, several steps are taken to maximize their reliability as well. These include employing a highly knowledgeable research staff who examine each judgement and sentence for conformity with the SRA, a custom data-entry program that does the same, and the use of monthly double-entry reliability checks to identify and correct any problems.

changes in the kinds of cases and changes in charging practices.¹² While both may change over time, it is unlikely that changes in the types of offenses referred for prosecution would follow the specific patterns predicted and observed here. We also assume that cases are declined, dismissed, or acquitted primarily for evidentiary reasons and that any error introduced by the loss of those cases should be uncorrelated with the sentencing reforms that are our focus.

We limit the data for the present analysis to "first-time offenders," those with no prior felony convictions known at the time of sentencing ($N = 15,489$). This is convenient for two reasons. First, the elimination of the FTOW (the first sentencing reform) only affects cases where there are no prior felonies. Second, it partially controls for the possibility that changes over time in convictions and sentences are due to differences in the criminal histories of those convicted of selling drugs. The conclusions and generalizations drawn from the analyses are limited accordingly.

Data included in this analysis are the primary (i.e., most serious) offense for which each offender was convicted (delivery, possession, or conspiracy to delivery), the presence of multiple charges at conviction, whether the sentence included any sentence enhancement, the type and length of sentence ordered (FTOW, prison vs. nonprison, months of confinement), the method of conviction (guilty plea vs. trial), the offender's sex, race/ethnicity, and the fiscal year (referred to simply as "year") in which the offender was convicted and sentenced. Table 1 presents descriptive statistics for each item.

ANALYSES AND RESULTS

In order to test the hypotheses outlined above, the analysis treats the *primary offense at conviction*, conviction on *multiple counts*, *sentence enhancements applied*, *type of sentence*, and *length of confinement ordered* as dependent variables, with *year* and *mode of conviction* as the main independent variables of interest. We use bivariate and descriptive analyses of the dependent variables, aggregated by year and by mode of conviction, as well as logistic regressions testing whether or not the charging and sentencing outcomes in individual cases are related to the year of sentencing, controlling for mode of conviction, sex, and race/ethnicity. Given the complexity of the changes we are examining, the descriptive analyses of changes over time in charging and sentencing outcomes are

¹² One way to control for the types of cases referred to prosecutors would be through the use of aggregate-level arrest data. Unfortunately, due to inconsistencies in law enforcement agencies' reporting practices in Washington, that is not currently possible.

TABLE 1
DESCRIPTIVE STATISTICS, 1986-1995^a

	Frequency	Mean/ Proportion of Total
Conviction charges:		
Primary offense:		
Possession (0 = no, 1 = yes)	9,190	.593
Delivery (0 = no, 1 = yes)	5,232	.338
Conspiracy to deliver (0 = no, 1 = yes)	1,067	.069
Total cases	15,489	
Conviction on multiple counts (delivery only)	1,198	.229
Conviction with sentence enhancements (delivery only)	246	.047
Sentencing outcomes:		
Mean sentence length, in months	8.60
Prison sentence ordered	4,723	.305
First-time offender waiver (all cases)	3,809	.246
First-time offender waiver (delivery only)	507	.097
Mode of conviction:		
Guilty plea	13,953	.901
Trial	1,536	.099
Offender characteristics:		
White	8,583	.554
Black	3,013	.195
Hispanic	3,594	.232
Other race/ethnicity	344	.022
Male	12,353	.798
Female	3,136	.202
Year of conviction, total cases:		
1986	456	.029
1987	795	.051
1988	1,346	.087
1989	2,127	.137
1990	2,085	.135
1991	1,714	.111
1992	1,784	.115
1993	1,862	.120
1994	1,707	.110
1995	1,613	.104

^a All variables are dichotomous (0 = no, 1 = yes), except mean sentence length.

more easily interpreted. We limit our discussion to those analyses that are consistent with the findings of the logistic regression analyses (for regression results, see tables A1–A6 in the appendix). Furthermore, those changes, which are our primary interest, are virtually unaffected by the inclusion of sex, race/ethnicity, and type of conviction in the logistic regression models.¹³

Analyses of Sentencing Outcomes

We begin by examining the impact of sentencing reforms on the severity of sentences. Changes in the severity of sentencing are directly relevant to hypotheses 1–5. Specifically, we examine the severity of sentences, by year, for: (a) offenders convicted of delivery (hypothesis 1), (b) for all offenders (hypothesis 3), and (c) for all offenders by mode of conviction (hypotheses 2, 4, and 5). Table 2 presents the percentage of offenders sentenced to prison and the average sentence length ordered, both for offenders convicted of delivery and for all drug offenders, by year.¹⁴ In addition, table 2 presents the percentage of FTOW sentences for offenders convicted of delivery.

The formal legal model predicts that the severity of sentencing will change in accordance with legal reforms (hypothesis 1). The data in table 2 show that the severity of sentences for offenders convicted of delivery increased substantially during the period studied. The FTOW was ordered in about 65% of delivery convictions in 1986, in only 6% in 1989, and virtually zero by 1990. Consequently, the percentage of prison sentences for deliveries increased in similar proportion, from 24% in 1986 to 88% in 1989, leveling off at about 93% (+/–1%) from 1991 to 1995. The average sentence length for delivery cases increased similarly, doubling from about six months in 1986 to 13 months in 1989 (precisely the midpoint of the presumptive range). Average sentences then jump to 23 months in 1990, and continue to climb, peaking at nearly 29 months in

¹³ Offenders' sex and race/ethnicity are significantly related to a number of the dependent variables examined, but because this may reflect unmeasured differences in actual offending, we do not interpret those findings as necessarily evidence of disparate treatment by the courts. However, to the extent that offending behavior or enforcement is related to sex or race/ethnicity, controlling for those offender characteristics may control, in part, for changes over time in the composition of cases referred for prosecution.

¹⁴ The elimination of the FTOW for drug dealers became effective in 1988. The reforms increasing presumptive sentence lengths for dealers all became effective in 1990. While the *Mendoza* ruling came late in 1992, it was not immediately applied in sentencing of conspiracy. Therefore, we treat 1993 as the first year in which the *Mendoza* rule would be followed routinely.

TABLE 2

SENTENCING OUTCOMES FOR ALL OFFENDERS AND FOR DELIVERIES ONLY

	ALL OFFENDERS			DELIVERIES ONLY			
	Total Cases	Prison Sentences (% of total)	Average Sentence (Months)	Total Cases	FTOW Sentences (% of total)	Prison Sentences (% of total)	Average Sentence (Months)
1986	456	50 (11.0)	3.3	203	131 (64.5)	48 (23.6)	6.3
1987	795	137 (17.2)	3.8	310	159 (51.3)	133 (42.9)	8.3
1988 ^a	1,346	197 (14.6)	3.4	398	169 (42.5)	196 (49.2)	8.9
1989	2,127	465 (21.9)	4.2	526	31 (5.9)	460 (87.5)	13.3
1990 ^b	2,085	617 (29.6)	8.1	659	6 (.9)	599 (90.9)	22.6
1991	1,714	679 (39.6)	11.4	661	3 (.5)	609 (92.1)	25.9
1992	1,784	701 (39.3)	12.4	690	3 (.4)	652 (94.5)	28.7
1993 ^c	1,862	693 (37.2)	10.7	663	3 (.5)	615 (92.8)	25.6
1994	1,707	643 (37.7)	11.4	627	0 (.0)	593 (94.6)	26.9
1995	1,613	541 (33.5)	10.6	495	2 (.4)	463 (93.5)	27.9
Total	15,489	4,723 (30.5)	8.6	5,232	507 (9.7)	4,368 (83.5)	21.8

^a Offenders convicted of delivery no longer eligible for FTOW.^b Seriousness level increased, scoring rules changed, and enhancements added for delivery cases.^c Conspiracy cases sentenced as unranked offenses (0-12 months).

1992. These findings that the severity of sentences did change in accordance with changes in the sentencing laws provide initial support for formal legal hypothesis 1.¹⁵

The substantive rationality model predicts that courtroom workgroups will adapt to reforms in ways that maintain established levels of punishment (hypothesis 3). Table 2 shows that, among all offenders, the mean length of sentence changed little from 1986 to 1989, increasing by only one month (from 3.3 to 4.2). Beginning in 1990, however, the average sentence increased substantially and had tripled by 1992 to 12.4 months, before decreasing to about 11 months in 1993 to 1995. The increases in overall sentence severity after 1990 clearly resulted from dramatic increases in sentences for offenders convicted of delivery. The overall prevalence of prison sentences followed a similar pattern.¹⁶ While it appears that courtroom workgroups may have largely neutralized the effect of losing the FTOW as a sentencing option, they clearly did not neutralize the 1990 reforms. This evidence leads us to reject hypothesis 3.

In contrast to the formal legal and substantive rationality models, the organizational efficiency model specifically predicts that the impact of sentencing reforms on sentences will be contingent on guilty pleas (hypothesis 5). Figures 1 and 2 present changes over time in the overall severity of sentencing by mode of conviction. The data are unequivocal—offenders who plead guilty are consistently sentenced less severely than offenders convicted at trial. Although this is true in each year, the magnitude of the disparity increases over time. While elimination of the FTOW had little impact on the sentences of offenders who pled guilty, it increased the severity of sentences for those convicted at trial. Sentence lengths increased again, and substantially, for *both* groups beginning in 1990. That increase, however, was much larger for offenders convicted at trial than for those who pled guilty. From 1990 to 1995, offenders who took their cases to trial received sentences averaging more than one year longer than offenders who pled guilty. This evidence provides strong support for hypothesis 5 and leads us to reject hypotheses 2 and 4.

¹⁵ Judges did not, however, entirely relinquish their discretion in sentencing. Analyses not shown here find that most of the delivery cases between 1990 and 1995 that were *not* sentenced to prison received *mitigated exceptional sentences*, which judges used in about 8% of all delivery cases after 1989. Aggravated exceptional sentences were far less common. In this way, judges have lessened, slightly, the overall impact of the reforms that otherwise dramatically increased sentences for those convicted of delivery of heroin or cocaine.

¹⁶ We look at both changes in sentences to prison and changes in average sentence length because, while they are closely related, those decisions may be affected by different factors. We find, however, that the two measures of severity yield essentially the same results, so we limit the remainder of our discussion to changes in average sentence length.

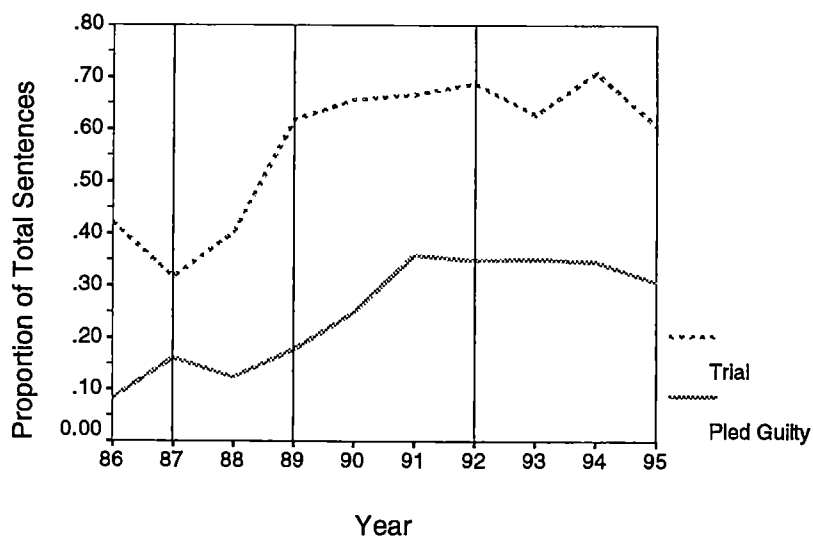


FIG. 1.—Prison sentences by mode of conviction ($N = 15,489$)

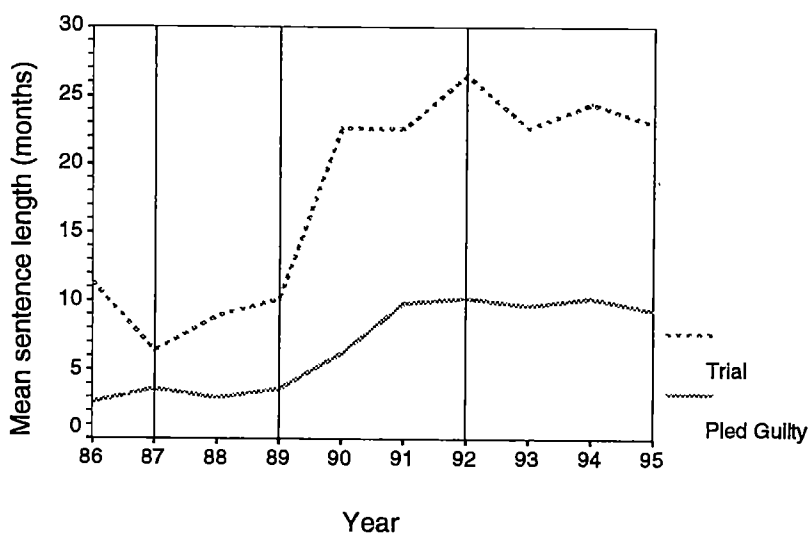


FIG. 2.—Length of sentence by mode of conviction ($N = 15,489$)

In sum, the analysis of sentencing outcomes over time provides partial support for the formal legal model (hypothesis 1) but leads us to reject hypothesis 2. We find contradictory evidence regarding substantive rationality hypothesis 3, leading us to reject it and hypothesis 4. These findings are consistent with organizational efficiency hypothesis 5 and thus provide initial support for that model. Next, we examine charging outcomes and whether reforms resulted in changes in charging practices that may have produced these changes in sentencing outcomes.

Analyses of Charging Outcomes

The analysis of changes in the severity of charging outcomes is critical to each of our theoretical models and to the more general test of the hydraulic displacement thesis. In this section, we examine changes in the severity of charges at conviction, first for all offenders and for offenders convicted of delivery (hypotheses 6 and 8) and second by mode of conviction (hypotheses 7 and 9–12).

The formal legal model predicts that changes over time in the severity of conviction charges will be independent of changes in the law (hypothesis 6). Table 3 presents the total number and percentages of convictions for narcotic violations by primary offense and year. The percentages of drug offenders convicted of *delivery*, *conspiracy to deliver*, and *possession* of narcotics vary substantially during this period ($\chi^2 = 1,200$; $df = 18$; $P < .01$). Furthermore, these changes do not appear to be random fluctuations. Rather, distinct patterns of changes over time corresponding closely with changes in the law are readily apparent in figure 3, leading us to reject hypothesis 6.

The substantive rationality model predicts that the severity of charges will decrease as reforms increase the sentences for delivery (hypothesis 8). The specific changes apparent in table 3 and figure 3 allow us to address this hypothesis. From 1986 to 1989, there was a marked increase in the percentage of convictions that were for simple possession (54% to 72%) and a corresponding decrease in convictions for delivery of narcotics (44% to 25%). These changes are consistent with hypothesis 8. This pattern *reverses* direction in the next two years, however, so that by 1991 the proportions of possession and delivery convictions are similar to those of 1987. The hypotheses derived from the substantive rationality model clearly would not predict such a reversal. The number of conspiracy convictions was trivial in the first five years, but it increased to about 4% of all convictions in 1991. These proportions do not change in 1992, but from 1993 to 1995, the prevalence of conspiracy convictions increases steadily to a substantial 21% of all convictions—a fivefold increase in only three years. Both delivery and possession convictions decrease slightly during

TABLE 3
CONVICTIONS BY OFFENSE AND YEAR

Year	Possession (%)	Conspiracy (%)	Delivery (%)	Total (%)
1986	245 (53.7)	8 (1.8)	203 (44.5)	456 (100)
1987	474 (59.6)	11 (1.4)	310 (39.0)	795 (100)
1988 ^a	924 (68.6)	24 (1.8)	398 (29.6)	1,346 (100)
1989	1,539 (72.4)	62 (2.9)	526 (24.7)	2,127 (100)
1990 ^b	1,382 (66.3)	44 (2.1)	659 (31.6)	2,085 (100)
1991	980 (57.2)	73 (4.3)	661 (38.7)	1,714 (100)
1992	1,017 (57.0)	77 (4.3)	690 (38.7)	1,784 (100)
1993 ^c	1,028 (55.2)	171 (9.2)	663 (35.6)	1,862 (100)
1994	822 (48.2)	258 (15.1)	627 (36.7)	1,707 (100)
1995	779 (48.3)	339 (21.0)	495 (30.7)	1,613 (100)
Total	9,190 (59.3)	1,067 (6.9)	5,232 (33.8)	15,489 (100)

^a Offenders convicted of delivery no longer eligible for FTOW.

^b Seriousness level increased, scoring rules changed, and enhancements added for delivery cases.

^c Conspiracy cases sentenced as unranked offenses (0–12 months).

this last period, though, suggesting that prosecutors may have used conspiracy charges in some cases to offset the lengthy sentences mandated for offenders charged with delivery. This substitution, however, did little to counter the increase in delivery convictions in the 1990s.

Conviction of delivery, however, is only one measure of charge severity. Prosecutors also control the number of counts and the application of sentence enhancements, which may make as much or more difference in the resulting sentence as would a reduction to possession. The substantive rationality model would predict that, if prosecutors were not reducing the severity of the primary charge, they would have at least mitigated the effects of the reforms by charging fewer offenders with multiple counts and foregoing the use of enhancements. In that case, offenders convicted of delivery would still go to prison, but the impact of the 1990 changes might be minimized.

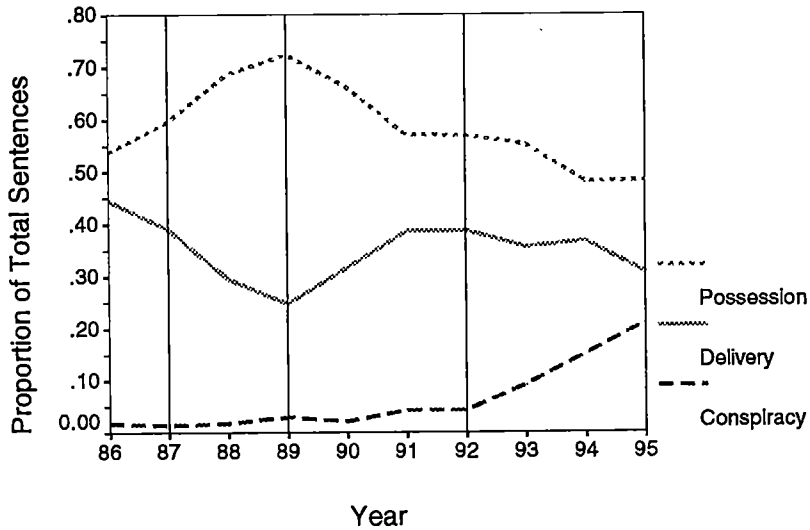


FIG. 3.—Conviction by type of offense, all cases ($N = 15,489$)

To test this possibility, we examine changes over time in two other measures of charge severity among delivery convictions: convictions on *multiple counts* and cases receiving *sentence enhancements*. Figure 4 shows the changes over time for these two measures.¹⁷ The prevalence of convictions on multiple counts decreases sharply from 32% of all cases in 1987 to about 14% in 1989, but *increases* again beginning in 1990 (when hypothesis 8 would predict a decrease), and fluctuates between about 20% and 25% of all cases thereafter. Sentence enhancements are much less prevalent but follow a very similar pattern, increasing from about 1% to about 4% in 1990 and doubling again to about 8% in 1992. These additional measures, then, do not help to explain why, from 1990 to 1992, delivery convictions became more common. While charging trends between 1987 and 1989 are consistent with hypothesis 8, the fact that, by each of our measures, the severity of charging *increased* subsequent to the 1990 reforms leads us to reject that hypothesis.

¹⁷ The number of counts is dichotomized because very few cases are charged with more than two. Recall that these sentence enhancements are *mandatory*, based on a "finding of fact" that the offender either possessed a deadly weapon at the time of the offense or that the offense was committed in a "protected zone." Prosecutors determine whether that information is presented in court, so these are essentially enhanced *charges*. Because deadly weapon enhancements were used rarely, we examine the use of *any* enhancements.

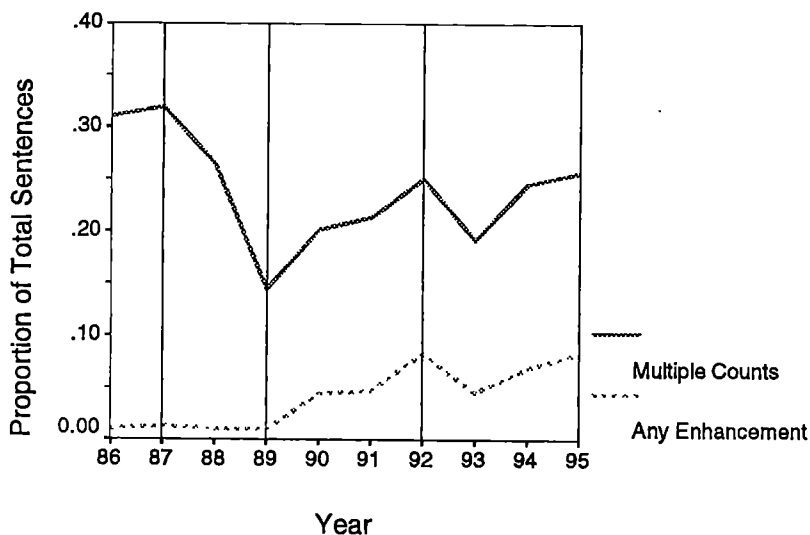


FIG. 4.—Deliveries with multiple counts or enhancements ($N = 5,232$)

The final stage of our analysis examines whether the severity of conviction charges, and changes in their severity, are contingent upon mode of conviction, as predicted by the organizational efficiency model (hypotheses 10–12), or not, as predicted by the formal legal and substantive rationality models (hypotheses 7 and 9, respectively). Figures 5 and 6 present the proportions of cases by offense, year, and mode of conviction.

Again, the data by mode of conviction are unequivocal. In every year, and by each of our measures, offenders who plead guilty are consistently convicted of less serious charges. On average, over the 10 years studied, 68% of offenders convicted at trial are convicted of delivery of narcotics, compared to 30% of those who plead guilty. Even more striking, of 1,067 convictions of conspiracy or other anticipatory offenses, *almost none* ($N = 13$; .8%) resulted from a trial, suggesting that anticipatory offenses are used almost exclusively as a bargaining chip in plea bargaining and simply do not indicate factual characteristics of offenses. Mode of conviction is also significantly related to conviction on multiple counts and findings of fact that require sentence enhancements. In tables not shown, offenders convicted at trial were, overall, twice as likely to be convicted of multiple counts (22%) than offenders who pled guilty ($\chi^2 = 186.36$; $df = 2$; $P < .01$), and from 1990 to 1995, were approximately *eight times* more likely to receive a sentence enhancement (16% of those who went to trial versus 2% of those who pled guilty; $\chi^2 = 159.96$; $df = 2$; $P < .01$). These differences are clearly consistent with, and offer strong support for, orga-

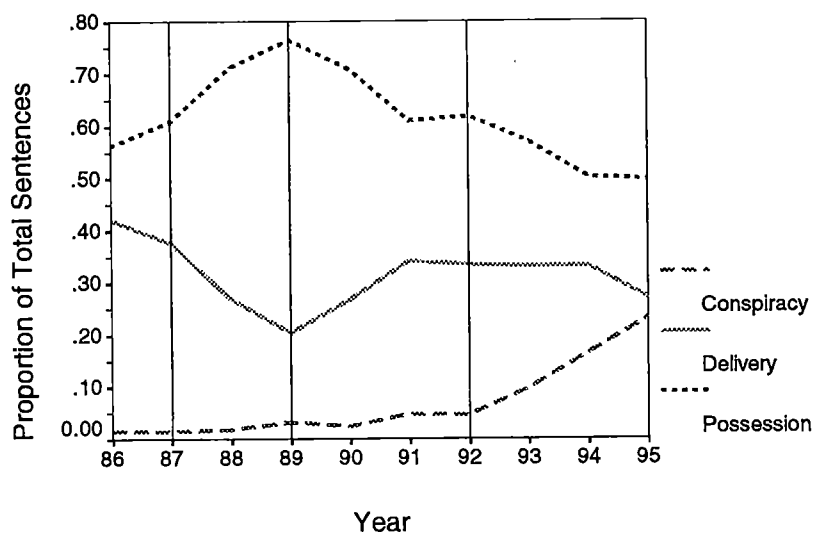


FIG. 5.—Convictions by offense, pled guilty only ($N = 13,953$)

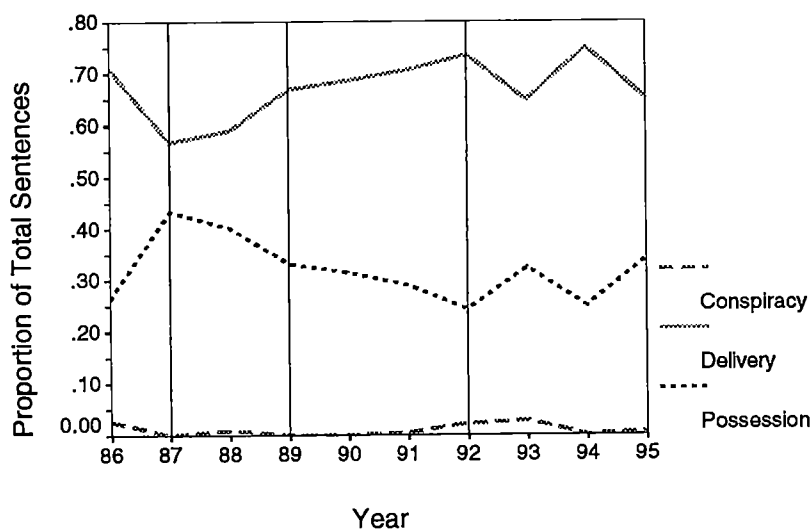


FIG. 6.—Convictions by offense, trials only ($N = 1,536$)

nizational efficiency hypothesis 10, and they lead us to reject hypotheses 7 and 9.

Perhaps the most significant finding in our analysis, however, is that the *changes over time* in the severity of charges are *entirely contingent upon the mode of conviction* (see figs. 5 and 6). The pattern of changes for cases that pled guilty is virtually identical to that described above for all cases, but the changes are even more pronounced. The prevalence of possession convictions, for example, increases from 56% in 1986 to 76% in 1989, while delivery convictions decrease by half; in 1990, the pattern reverses direction, with possession cases decreasing to about 60% of all guilty pleas and deliveries increasing to about one third. Conspiracy convictions account for 5% of all cases pled guilty in 1991 and 1992, increasing to 23% by 1995.

Most importantly, changes over time in the primary offense among cases convicted at trial are *unrelated* to sentencing reforms. Among those cases, the only apparent pattern is a gradual increase over time in the prevalence of delivery convictions, from 57% in 1987 to around 70% from 1991 to 1995. Possession convictions necessarily decrease over time in virtually identical proportions. Furthermore, the changes in conviction on multiple counts and the use of sentence enhancements (shown above in fig. 4) are similarly contingent upon the mode of conviction: the decrease in convictions on multiple counts only occurred for offenders who pled guilty, and the increase in sentence enhancements in 1990 only affected cases convicted at trial. These findings provide strong support for organizational efficiency hypotheses 11 and 12.

It is important to note, as a final observation, that these reforms also affected the rate at which offenders pled guilty. While the plea rate did not change significantly until 1990, the percentage of convictions obtained through guilty pleas decreased steadily from about 92% in 1987 to a low of 87% in 1992 (see fig. 7). This may reflect a growing unwillingness on the part of defendants to plead guilty to increasingly severe charges that entailed long prison sentences. The plea rate jumps again to 92%, though, in 1993, following the *Mendoza* ruling. This suggests that prosecutors were able to use charge reductions to conspiracy in order to bring the plea rate back to the level at which it had been prior to 1990.

The analyses of charging and sentencing by mode of conviction indicate, quite convincingly, that the impact of the sentencing reforms depends upon whether offenders plead guilty or are convicted in a trial. The elimination of the FTOW had little effect on offenders who pled guilty, while significantly increasing the severity of sentences for those convicted at trial. The 1990 reforms resulted in more severe sentencing for both groups of offenders but had a much greater impact on those who were convicted at trial. These differences in sentencing outcomes reflect the

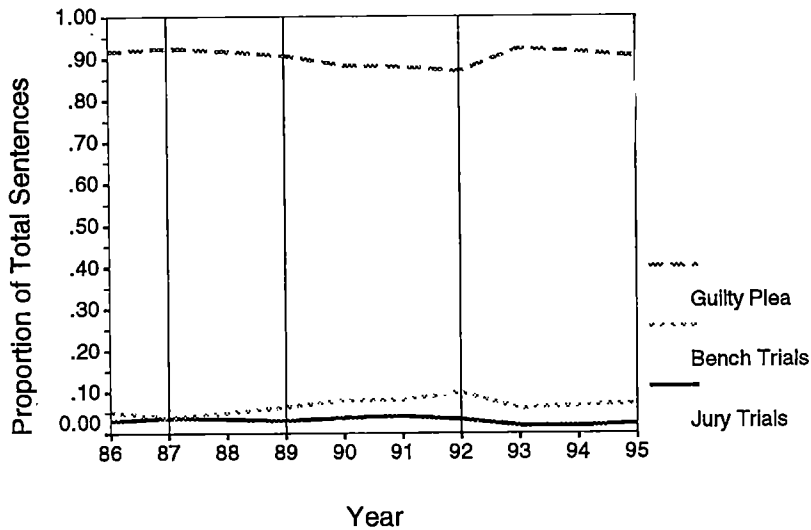


FIG. 7.—Mode of conviction by year ($N = 15,489$)

severity of charges for which the two groups were convicted. While charge severity increased following the 1990 reforms for both groups of offenders, the protected-zone and deadly weapon enhancements were reserved almost exclusively for offenders convicted at trial. Conversely, the sharp increase in convictions on the reduced charge of conspiracy to deliver narcotics was reserved for those who pled guilty. These results are clearly consistent with the hypotheses derived from the organizational efficiency model, while largely refuting those derived from the formal legal and substantive rationality models. While we cannot conclude from this evidence that formal rational and substantively rational concerns are irrelevant, we do conclude that the need for organizational efficiency provides a more parsimonious interpretation of the changes we observed.

DISCUSSION AND CONCLUSION

This article began with the questions of whether and how organizational processes and the use of discretion affect the implementation of sentencing reforms. Those questions were driven, in part, by the common criticism that the goals of sentencing guidelines will be subverted by the hydraulic displacement of discretion and by observations such as Savelsberg's: "sentencing guidelines disregard a basic sociological fact of modern organizations. Sentencing, like all decision making, is not an isolated event but an element in a chain of interconnected events. The intervention in

sentencing decisions as one element of the criminal justice system results in possibly neutralizing reactions in other parts of that system" (1992, p. 1372).

Our review of the theoretical literature suggested three alternative models of the sentencing process that, while not necessarily incompatible, suggest different types of goals that are likely to influence not only the application of existing laws, but also courts' adaptations when laws change. The sentencing process may be driven by: formal legal rationality, which is oriented toward the law; substantive rationality, which is oriented toward sentencing outcomes; or organizational efficiency, which is oriented toward the process itself.

Clearly, judges in Washington State have complied with the sentencing reforms, as evidenced by their discontinuing use of the FTOW for drug dealers and their ultimately sentencing 93% of those offenders to prison in accordance with the increased presumptive ranges. However, the bulk of the evidence presented here leads us to reject the formal legal hypothesis that organizational processes have little or no effect on the implementation of sentencing reforms. The severity of charges at conviction changed significantly following each change in the law, which suggests the manipulation of charges (and subsequent sentences) rather than a strict application of charges to the crimes committed. While the severity of sentencing for delivery of narcotics increased following the 1988 and 1990 reforms, as predicted by the formal legal model, the overall impact of those reforms is substantially greater for offenders convicted at trial than for those who pled guilty.

Some of the evidence is consistent with the argument based on substantive rationality that, under sentencing guidelines, courtroom workgroups will manipulate charges so that, overall, a consistent level of punishment is maintained for a group of offenders. The findings that the severity of charges decreased sharply, and that the use of prison sentences and the average lengths of sentences did not change appreciably when the FTOW was eliminated, would seem to support the general hypothesis. However, that adaptation was entirely contingent upon offenders having pled guilty. Furthermore, the substantive rationality model, as we understand it, *cannot* explain the sharp increase in severity of punishment that began in 1990 and that affected those who pled guilty as well as those convicted at trial. By continuing, or even increasing, the use of charge reductions from delivery to simple possession or anticipatory offenses, courtroom workgroups clearly could have maintained a stable rate of punishment. They did not.

We interpret the evidence here, both on the whole and in specific instances, to be consistent with predictions based on the operational goal of organizational efficiency. In each instance of reform to the sentencing

laws, the consequences for offenders are contingent upon whether they pled guilty or were convicted in a trial and are consistently more severe for the latter group. This takes place irrespective of the substantive severity of the resulting punishments, which changed substantially even for those who pled guilty. Finally, the distinct changes in the composition of offenses at conviction—the initial increase in possession convictions, their subsequent rapid decline, and later the dramatic increase in convictions on conspiracy to deliver narcotics—are consistent with the general prediction of the efficiency model, that charging practices will change as options for inducing guilty pleas change. Most importantly, those changes are *entirely* contingent upon offenders having pled guilty.

While the evidence presented here appears to be strong, there are several limitations to this study that require comment. The first, and most serious, of these is that like most studies of the impact of sentencing reform we are limited to conviction data. We partially overcome that limitation by focusing on changes over time at the aggregate level. The validity of our analysis is therefore contingent upon the assumption that the characteristics of cases referred for prosecution are, if not constant over time, at least uncorrelated with sentencing reforms.

We cannot verify that assumption, but it is unlikely that the changes we observe in conviction charges simply reflect changes in the types of cases referred. While the nature of drug offending and drug law enforcement probably *did* change during the period studied, it is not at all clear how such changes would lead to the reversing pattern in types of conviction charges that took place. Most importantly, if that were the case, we would expect to see those changes reflected in trial convictions as well as cases that pled guilty. Furthermore, a previous study of drug-offender sentencing in Washington State (Engen and Steiger 1997) offers compelling evidence that the increase in convictions of *conspiracy* to deliver narcotics reflects a change in charging practices, rather than a change in offending. In that study, interviews with 27 judges, prosecutors, and public defenders in the four largest counties revealed—*explicitly and unanimously*—that these conspiracy convictions are in fact plea bargains in cases that could have been charged as delivery.

A second limitation of this study is that it focuses on a specific population, first-time drug offenders. Drug offending is a particularly politically charged crime, subject to changing public perceptions and a more “tightly-coupled” response by the criminal justice system (Peterson and Hagan 1984; Hagan 1989). Given the politicization of drug offender sentencing in Washington State in the late 1980s, it is possible that the changes we observed were influenced by specific external political forces of which we are unaware, in addition to more generic organizational processes. For instance, the fact that use of the FTOW for drug dealers decreased sub-

stantially and that prison sentences were increasing in frequency even *before* 1988 suggests that other forces may have contributed to the FTOW's demise. Future research should examine the implementation of reforms affecting other types of offenses and offenders.

Third, the adaptations observed here may be unique, due either to the nature of Washington's sentencing guidelines, or the specific reforms that took place. Washington's guidelines are relatively "tight," providing less judicial discretion than the more "loose" guidelines adopted in some states (Tonry 1996). As a result, charge manipulation in response to reforms may be more likely in Washington than in other states (Savelsberg 1992; Ulmer 1997). Also, the specific changes to the sentencing laws, which were unique and in some cases extreme, may not be representative of reforms elsewhere. Hagan, though, calls for research in "contexts where the surrounding political environment has mandated departures from normal criminal justice operations," arguing that "we learn a great deal about system operations in those circumstances" (1989, p. 130). The reforms that we examine here seem to fit that description quite well, so may add value our analysis. To the extent that the issues we examine are generic, our conclusions regarding the relative influence of formal rational, substantive, and organizational concerns may generalize to other legal contexts, even if the mechanisms by which those are achieved do not. Future research should, in any case, examine adaptations to changes in sentencing guidelines in other states.

Fourth, researchers have demonstrated that sentencing under guidelines, and courtroom workgroups' use of guidelines, can vary among courts under the same laws (e.g., Miethe 1987; Dixon 1995; Kramer and Ulmer 1996; Albonetti 1997; Engen and Steiger 1997; Ulmer 1997). Our analysis is appropriate to the question of whether societal (in this case, state) mandates to change sentencing achieve their intended goals, but it does not speak to organizational differences in adaptations to sentencing reforms. Our analysis does not take into account local differences in organizational structure, caseload pressures, or resources (e.g., jail space) that might reasonably affect courts' adaptations to changes in the laws. Preliminary analyses examining differences across counties indicate that the patterns observed at the state level are typical, but future research should compare adaptations to changes in sentencing laws across organizational contexts, examining those factors that are beyond the scope of this analysis.

In spite of these limitations, our analyses suggest that it is premature to conclude that sentencing guidelines have achieved their goal of uniform sentencing for similarly situated offenders. On the contrary, the *hydraulic displacement* of discretion appears to be a very real phenomenon. Organizational adaptations subsequent to each of the sentencing reforms exam-

ined here appear to have produced very different sentences for similar offenders. One implication of this finding is that if those charging decisions and adaptations are related to offenders' status characteristics, such as sex, race, or ethnicity, the effects of those characteristics on punishment may be underestimated by research that focuses exclusively on the sentencing stage. More importantly, the conclusion that sentencing guidelines have reduced inequality in criminal justice may be plainly wrong. Research on inequality in social control *must* address this possibility. A second implication is that some of the reforms introduced in Washington may have resulted in *less uniform* sentencing for similar offenders than existed prior to the reforms, and thus may have worked counter to the goal of uniformity. Research examining that possibility is critical if policy makers are to make informed decisions.

Clearly, however, sentencing guidelines, and the reforms that took place in Washington State, are not irrelevant to punishment. The punishment of drug offenders is shaped by and reflective of those laws, but sentencing reforms do not directly determine sentencing outcomes. Neither would we conclude that sentencing occurs irrespective of substantive concerns about appropriate sanctions for drug offenders. However, the bulk of the evidence presented here suggests that organizational needs overshadowed both the formal demands of sentencing reforms and substantively rational demands for appropriate punishment. It appears that a principal effect of sentencing reforms is to structure the ways in which courtroom workgroups use their discretion in charging and plea bargaining to encourage guilty pleas and thereby maintain organizational efficiency.

APPENDIX

TABLE A1

LOGISTIC REGRESSIONS PREDICTING CONVICTION CHARACTERISTICS FOR FIRST-TIME
NARCOTICS OFFENDERS

	Possession	Delivery	Conspiracy	Multiple Counts	Sentence Enhanced
Independent variable:					
Year:					
1987 (vs. 1986)	1.24	.81	.82	.93	...
1988 (vs. 1987)	1.53*	.63*	1.30	.63*	...
1989 (vs. 1988)	1.26*	.75*	1.52	.62*	...
1990 (vs. 1989)75*	1.38*	.75	1.23	3.71*
1991 (vs. 1990)66*	1.41*	1.98*	1.12	1.09
1992 (vs. 1991)	1.02	.96	1.06	1.09	1.79
1993 (vs. 1992)87	.96	2.08*	.82	.65
1994 (vs. 1993)76*	1.04	1.78*	1.22	1.45
1995 (vs. 1994)	1.03	.73*	1.54*	.90	1.07
Race/ethnicity:					
Black	1.03	.87*	1.60*	.66*	2.31*
Hispanic74*	.96	3.22*	1.43*	1.77*
Other	1.16	.83	1.17	.88	1.23
Sex (male = 1)69*	1.51*	.94	.92	1.22
Conviction:					
Jury trial59*	2.13*	.24*	1.40	5.45*
Bench trial17*	7.81*	.05*	2.91*	4.28*
Model statistics:					
-2LL (null)	20,929.54	19,812.21	7,767.79	11,259.85	1,984.40
-2LL (model)	19,636.29	18,567.94	6,535.50	10,917.85	1,733.13
χ^2	1,293.24	1,244.27	1,232.30	342	251.27

NOTE.—Estimated effects are exponents of the logistic regression coefficients (Exp[B]). For each model $P < .000$, with $df = 15$. The analysis predicting sentence enhancements is limited to offenders convicted of delivery of narcotics ($N = 5,232$). Also, because very few enhancements were used prior to 1990, years 1986–89 are combined in that analysis. Accordingly, the Exp(B) for 1990 represents the change in the odds of receiving an enhancement in 1990 compared to all previous years combined. For all other models, $N = 15,489$, and Exp(B) for each year represents the change in odds from the previous year.

* $P < .01$.

TABLE A2
LOGISTIC REGRESSIONS PREDICTING SENTENCE TYPE (PRISON; FTOW) FOR FIRST-
TIME NARCOTICS OFFENDERS

	ALL CASES		DELIVERY ONLY	
	Prison	FTOW	Prison	FTOW
Independent variable:				
Year:				
1987 (vs. 1986)	1.85*	.69*	2.91*	.53*
1988 (vs. 1987)80	.72*	1.27	.73
1989 (vs. 1988)	1.58*	.55*	7.19*	.09*
1990 (vs. 1989)	1.53*	.97	1.55	.14*
1991 (vs. 1990)	1.62*	.62*	1.12	.52
1992 (vs. 1991)96	.87	1.61	.91
1993 (vs. 1992)98	.88	.74	1.00
1994 (vs. 1993)	1.00	.86	1.38	.01
1995 (vs. 1994)80*	1.39*	.72	114.94
Race/ethnicity:				
Black94	.90	.85	.98
Hispanic	1.57*	.43*	5.58*	.16*
Other92	1.11	.86	.90
Sex (male = 1)	1.77*	.51*	2.47*	.44*
Conviction:				
Jury trial	1.92*	.79	1.01	.68
Bench trial	6.67*	.36*	1.86*	.34*
Model statistics:				
-2LL (null)	19,050.87	17,279.66	4,688.83	3,220.98
-2LL (model)	17,272.58	15,593.60	3,238.20	1,548.97
χ^2	1,778.28	1,686.06	1,450.63	1,772.00

NOTE.—Estimated effects are exponents of the logistic regression coefficients (Exp[B]). For each model $P < .000$, with $df = 15$.

* $P < .01$.

TABLE A3
OLS REGRESSIONS PREDICTING SENTENCE LENGTH (TOTAL
MONTHS) FOR FIRST-TIME NARCOTICS OFFENDERS

	All Cases	Delivery Only
Independent variable:		
Year:		
1987 (vs. 1986)87	2.40
1988 (vs. 1987)	-.48	.27
1989 (vs. 1988)35	3.12*
1990 (vs. 1989)	3.75*	9.15*
1991 (vs. 1990)	3.16*	3.04*
1992 (vs. 1991)85	3.08*
1993 (vs. 1992)	-1.14	-2.53*
1994 (vs. 1993)57	1.05
1995 (vs. 1994)	-.91	.39
Race/ethnicity:		
Black	-.38	-1.06
Hispanic	3.88*	6.03*
Other72	2.35
Sex (male = 1)	2.30*	2.06*
Conviction:		
Jury trial	4.80*	3.74*
Bench trial	15.24*	8.35*
Adjusted R^215	.24

NOTE.—Regression coefficients are unstandardized.

* $P < .01$.

TABLE A4

LOGISTIC REGRESSIONS PREDICTING THE PRIMARY OFFENSE AT CONVICTION FOR
FIRST-TIME NARCOTICS OFFENDERS WHO EITHER PLED GUILTY ($N = 13,953$) OR
WERE CONVICTED BY TRIAL ($N = 1,536$)

	POSSESSION		DELIVERY		CONSPIRACY	
	Plea	Trial	Plea	Trial	Plea	Trial
Year:						
1987 (vs. 1986)	1.20	2.23	.89	.52	.94	
1988 (vs. 1987)	1.59*	.86	.61*	1.11	1.24	
1989 (vs. 1988)	1.33*	.74	.70*	1.42	1.59	
1990 (vs. 1989)73*	.94	1.45*	1.06	.76	
1991 (vs. 1990)64*	.91	1.45*	1.08	1.96*	
1992 (vs. 1991)	1.04	.79	.96	1.15	1.00	
1993 (vs. 1992)83*	1.47	.99	.67	2.16*	
1994 (vs. 1993)76*	.68	1.00	1.66	1.84*	
1995 (vs. 1994)98	1.74	.74*	.56	1.54*	
Race/ethnicity:						
Black	1.08	.59*	.81*	1.69*	1.61*	
Hispanic78*	.53*	.88*	1.73*	3.21*	
Other	1.16	1.38	.83	.75	1.19	
Sex (male = 1)69*	.68	1.53*	1.49	.94	
Model statistics:						
-2LL (null)	18,468.3	1,901.6	17,045.2	1,921.9	7,471.5	
-2LL (model)	17,834.6	1,846.9	16,754.9	1,872.9	6,390.0	
χ^2	633.76	54.64	290.30	48.96	1,081.46	

NOTE.—Estimated effects are exponents of the logistic regression coefficients ($\text{Exp}[B]$). For each model $P < .000$, with $df = 13$. The analysis of anticipatory deliveries is limited to offenders who pled guilty because, in 10 years, only 13 offenders (0.8%) were convicted of those offenses in trials.

* $P < .01$.

TABLE A5

LOGISTIC REGRESSIONS PREDICTING MULTIPLE COUNTS AND SENTENCE
ENHANCEMENTS FOR FIRST-TIME NARCOTIC OFFENDERS WHO EITHER PLED GUILTY
OR WERE CONVICTED BY TRIAL

	MULTIPLE COUNTS		SENTENCE ENHANCED	
	Plea	Trial	Plea	Trial
Year:				
1987 (vs. 1986)86	1.52		
1988 (vs. 1987)65*	.53		
1989 (vs. 1988)66*	.46		
1990 (vs. 1989)	1.18	1.55	1.62	9.12*
1991 (vs. 1990)	1.16	.96	1.04	1.19
1992 (vs. 1991)	1.04	1.42	3.36*	1.06
1993 (vs. 1992)87	.64	.60	.64
1994 (vs. 1993)	1.21	1.40	1.25	1.90
1995 (vs. 1994)85	1.12	1.01	1.11
Race/ethnicity:				
Black63*	.92	2.24*	2.50*
Hispanic	1.27*	2.51*	1.36	2.44*
Other88	.78	1.49	.73
Sex (male = 1)89	1.43	1.15	1.35
Model statistics:				
No. of cases	13,953	1,536	4,185	1,047
-2LL (null)	9,523.3	1,603.5	1,081.9	773.7
-2LL (model)	9,383.8	1,525.4	1,021.8	694.9
χ^2	139.53	78.14	60.09	78.86

NOTE.—Estimated effects are exponents of the logistic regression coefficients (Exp[B]). For multiple counts models $P < .000$, with $df = 13$. For sentence-enhanced models, $P < .000$, with $df = 10$.

* $P < .01$.

TABLE A6

LOGISTIC REGRESSIONS PREDICTING PRISON SENTENCES AND OLS REGRESSIONS
PREDICTING TOTAL SENTENCE FOR FIRST-TIME NARCOTIC OFFENDERS WHO EITHER
PLED GUILTY ($N = 13,953$) OR WERE CONVICTED BY TRIAL ($N = 1,536$)

	PRISON SENTENCE		SENTENCE LENGTH	
	Plea	Trial	Plea	Trial
Year:				
1987 (vs. 1986)	2.23*	.60	1.16	-5.16
1988 (vs. 1987)74	1.54	-.62	3.01
1989 (vs. 1988)	1.50*	2.44*	.50	.81
1990 (vs. 1989)	1.59*	1.16	2.74*	12.12*
1991 (vs. 1990)	1.72*	1.03	3.61*	-.56
1992 (vs. 1991)95	1.13	.45	4.56
1993 (vs. 1992)	1.00	.78	-.67	-3.73
1994 (vs. 1993)97	1.46	.51	1.48
1995 (vs. 1994)83	.56	-.88	-2.32
Race/ethnicity:				
Black90	1.28	-.49	1.65
Hispanic	1.48*	2.38*	2.81*	12.33*
Other99	.46	.64	-.69
Sex (male = 1)	1.77*	1.86*	2.25*	4.48*
Model statistics:				
-2LL (null)	16,291.70	2,044.18		
-2LL (model)	15,420.16	1,912.63		
χ^2	871.54	131.55		
Adjusted R^208	.15

NOTE.—Estimated effects predicting prison sentence are exponents of the logistic regression coefficients (Exp[B]). OLS regression coefficients predicting sentence length are unstandardized. For each model $P < .000$, with $df = 13$.

* $P < .01$.

REFERENCES

- Albonetti, Celesta A. 1991. "An Integration of Theories to Explain Judicial Discretion." *Social Problems* 38:247-66.
- . 1992. "Charge Reduction: An Analysis of Prosecutorial Discretion in Burglary and Robbery Cases." *Journal of Quantitative Criminology* 8 (3): 317-33.
- . 1997. "Sentencing under the Federal Sentencing Guidelines: Effects of Defendant Characteristics, Guilty Pleas, and Departures on Sentence Outcomes for Drug Offenses, 1991-1992." *Law and Society Review* 31 (4): 789-822.
- Alschuler, Albert. 1978. "Sentencing Reform and Prosecutorial Power: A Critique of Recent Proposals for Fixed and Presumptive Sentencing." In *Determinate Sentencing: Reform or Regression*. Washington, D.C.: Government Printing Office.
- Balbus, Isaac. 1973. *The Dialectics of Legal Repression*. New York: Russell Sage.
- Boerner, David. 1995. "Sentencing Guidelines and Prosecutorial Discretion." *Judicature* 78 (4): 196-200.
- Bridges, George S., and Robert D. Crutchfield. 1988. "Law, Social Standing and Racial Disparities in Imprisonment." *Social Forces* 66 (3): 699-724.

- Daly, Kathleen. 1987. "Discrimination in the Criminal Courts: Family, Gender, and the Problem of Equal Treatment." *Social Forces* 66:152-75.
- . 1994. *Gender, Crime, and Punishment*. New Haven, Conn.: Yale University Press.
- Dixon, Jo. 1995. "The Organizational Context of Criminal Sentencing." *American Journal of Sociology* 100 (5): 1157-98.
- Eisenstein, James, and Herbert Jacob. 1977. *Felony Justice: An Organizational Analysis of Criminal Courts*. Boston: Little, Brown.
- Eisenstein, James, Roy Flemming, and Peter Nardulli. 1988. *The Contours of Justice: Communities and Their Courts*. Boston: Little, Brown.
- Engen, Rodney L., and John Steiger. 1997. "Trading Time for Treatment: Preliminary Evaluation of the Drug Offender Sentencing Alternative." Olympia: State of Washington Sentencing Guidelines Commission.
- Ewing, Sally. 1987. "Formal Justice and the Spirit of Capitalism: Max Weber's Sociology of Law." *Law and Society Review* 21 (3): 487-512.
- Frase, Richard S. 1993. "Implementing Commission-Based Sentencing Guidelines: The Lessons of the First Ten Years in Minnesota." *Cornell Journal of Law and Public Policy* 2:278-337.
- Hagan, John. 1989. "Why Is There So Little Criminal Justice Theory? Micro-Level Links between Organization and Power." *Journal of Research in Crime and Delinquency* 26 (2): 116-35.
- Knapp, Kay A. 1987. "Implementation of the Minnesota Guidelines: Can the Innovative Spirit Be Preserved?" Pp. 127-41 in *The Sentencing Commission and Its Guidelines*, edited by Andrew von Hirsch, Kay A. Knapp, and Michael Tonry. Boston: Northeastern University Press.
- Kramer, John, and Darrell Steffensmeier. 1993. "Race and Imprisonment Decisions." *Sociological Quarterly* 34 (2): 357-76.
- Kramer, John, and Jeffery J. Ulmer. 1996. "Sentencing Disparity and Departure from Guidelines." *Justice Quarterly* 13:81-106.
- Miethe, Terence D. 1987. "Charging and Plea Bargaining under Determinate Sentencing: An Investigation of the Hydraulic Displacement of Discretion." *Journal of Criminal Law and Criminology* 78 (1): 155-76.
- Miethe, Terence D., and Charles A. Moore. 1985. "Socioeconomic Disparities under Determinate Sentencing Systems: A Comparison of Pre-guideline and Post-guideline Practices in Minnesota." *Criminology* 23 (2): 337-63.
- Moore, Charles A., and Terence D. Miethe. 1986. "Regulated and Unregulated Sentencing Decisions: An Analysis of First-Year Practices under Minnesota's Felony Sentencing Guidelines." *Law and Society Review* 20:253-77.
- Nagel, Ilene H., and Stephen J. Schulhofer. 1992. "A Tale of Three Cities: An Empirical Study of Charging and Bargaining Practices under the Federal Sentencing Guidelines." *Southern California Law Review* 66:501-61.
- Nardulli, Peter F., Roy B. Flemming, and James Eisenstein. 1985. "Criminal Courts and Bureaucratic Justice: Concessions and Consensus in the Guilty Plea Process." *Journal of Criminal Law and Criminology* 76 (4): 1103-31.
- Packer, Herbert L. 1968. *The Limits of the Criminal Sanction*. Stanford, Calif.: Stanford University Press.
- Peterson, Ruth D., and John Hagan. 1984. "Changing Conceptions of Race: Toward an Account of Anomalous Findings of Sentencing Research." *American Sociological Review* 49:56-70.
- Rosett, Arthur, and Donald Cressey. 1976. *Justice by Consent*. Philadelphia: Lippincott.
- Savelsberg, Joachim. 1992. "Law That Does Not Fit Society: Sentencing Guidelines as a Neoclassical Reaction to the Dilemmas of Substantivized Law." *American Journal of Sociology* 97 (5): 1346-81.

- Steffensmeier, Darrell, John H. Kramer, and Cathy Streifel. 1993. "Gender and Imprisonment Decisions." *Criminology* 31:411–46.
- Steffensmeier, Darrell, Jeffery Ulmer, and John Kramer. 1998. "The Interaction of Race, Gender, and Age in Criminal Sentencing: The Punishment Cost of Being Young, Black, and Male." *Criminology* 36 (4): 763–98.
- Sudnow, David. 1965. "Normal Crimes: Sociological Features of the Penal Code." *Social Problems* 12 (4): 255–64.
- Tonry, Michael. 1996. *Sentencing Matters*. New York: Oxford University Press.
- Ulmer, Jeffery T. 1997. *Social Worlds of Sentencing: Court Communities under Sentencing Guidelines*. Albany: State University of New York Press.
- Ulmer, Jeffery, and John Kramer. 1996. "Court Communities under Sentencing Guidelines: Dilemmas of Formal Rationality and Sentencing Disparity." *Criminology* 34 (3): 383–407.
- Walker, Samuel. 1993. *Taming the System*. New York: Oxford University Press.
- Weber, Max. 1968. *Economy and Society*. New York: Bedminster Press.

Risk and Trust in Social Exchange: An Experimental Test of a Classical Proposition¹

Linda D. Molm, Nobuyuki Takahashi, and Gretchen Peterson
University of Arizona

The classical exchange theorists proposed that trust is more likely to develop between partners when exchange occurs without explicit negotiations or binding agreements. Under these conditions, the risk and uncertainty of exchange provide the opportunity for partners to demonstrate their trustworthiness. This study develops the theoretical implications of this proposition and conducts an experimental test that compares levels of both trust and commitment in two forms of direct exchange, *negotiated* and *reciprocal*. The results support the classical proposition, showing that reciprocal exchange produces stronger trust and affective commitment than negotiated exchange, and that behaviors signaling the partner's trustworthiness have greater impact on trust in reciprocal exchange.

Social exchange . . . involves favors that create diffuse future obligations, not precisely specified ones, and the nature of the return cannot be bargained about but must be left to the discretion of the one who makes it. . . . Since there is no way to assure an appropriate return for a favor, social exchange requires trusting others to discharge their obligations. (Blau 1964, pp. 93–94)

For Blau and most of the other classical exchange theorists, “social” exchange referred to reciprocal acts of benefit, in which individuals offer help, advice, approval, and so forth to one another without negotiation of terms and without knowledge of whether or when the other will reciprocate. These exchanges necessarily entail uncertainty and risk: the other

¹ This research was supported by a grant from the National Science Foundation (SBR-9514911) to the first author. We gratefully acknowledge their support and the able assistance of Alex Marek and Tara Ripley with the research. We also thank the *AJS* reviewers for helpful comments on earlier drafts. The contributions of the second and third authors are equal. Direct correspondence to Linda Molm, Department of Sociology, University of Arizona, Social Science Building, Room 400, Tucson, Arizona 85721. E-mail: molml@u.arizona.edu

might never reciprocate or might do so very minimally. But rather than endangering social exchange, Blau, Lévi-Strauss, and others argued that the risk and uncertainty inherent in these reciprocal exchanges are essential for the development of trust and commitment. Another's trustworthiness can be demonstrated only when exchange occurs without the explicit "quid pro quo" of transactions that stipulate returns, and without the assurance of binding agreements (Blau 1964; Ekeh 1974; Lévi-Strauss 1969).

Contemporary theorists have continued to emphasize the link between risk and trust (e.g., Gambetta 1988; Kelley and Thibaut 1978; Luhmann 1979), and some, like Coleman (1990), have distinguished between reciprocal and negotiated exchange on these dimensions. But despite increasing recognition of the importance of trust in society (and concerns that it may be declining, e.g., Putnam [1995]), empirical examinations of trust in social relations are rare (see Kollock [1994] for an exception). Consequently, the prediction of the classical theorists—that trust is more likely to develop in reciprocal exchanges than in negotiated exchanges—has never been tested. Instead, studies of negotiated exchange, in which actors bargain over the terms of strictly binding agreements, have increasingly dominated social exchange research (e.g., Cook et al. 1983; Markovsky et al. 1993). The terms of these exchanges are known in advance and guaranteed, and consequently trust is unnecessary (Bonacich 1995).

In this study, we report the results of an experiment testing the classical prediction that trust is more likely to develop in reciprocal than in negotiated exchanges. This work is part of a larger research program, the first to compare these two forms of exchange and their effects on power, trust, and related processes.

Our analysis draws on Yamagishi and Yamagishi's (1994) distinction between *trust*, expectations of benign behavior based on inferences about a partner's personal traits and intentions, and *assurance*, expectations that are based, instead, on knowledge of an incentive structure that encourages benign behavior. We propose that negotiated exchanges with binding agreements provide assurance, while reciprocal exchanges enable trust. Whether trust actually develops depends on the partner's behavior and the information it conveys about the partner's trustworthiness. In support of these predictions, our results show that reciprocal acts of individual giving produce significantly higher levels of trust than the joint negotiation of binding agreements, even when exchanges of equivalent value, in equivalent structures of power and opportunity, are compared. And, when the risk of reciprocal exchange provides the opportunity to demonstrate trustworthiness, actors' behaviors—their commitments to one another and the equality or inequality of their exchanges—strongly influence the level of trust that develops.

We link the development of trust to recent work on commitment in exchange relations (Kollock 1994; Lawler and Yoon 1996), both *behavioral commitment* (repeated exchanges between the same actors) and *affective commitment* (feelings of positive affect and commitment toward the partner). We propose that trust is one aspect of a broader nexus of feelings toward the partner, which also includes affective commitment. We show that reciprocal exchanges produce both higher levels of trust and stronger feelings of affective attachment and commitment to the partner than negotiated exchanges. Behavioral commitments, on the other hand, are primarily influenced by the structure of power. Power affects trust and affective commitment through these behaviors, which build trust by signaling the other's trustworthiness and reducing the inequality of exchange.

The signaling effects of behavioral commitments are largely restricted to reciprocal exchanges, however. Even when negotiated exchanges occur within a continuing relation, rather than on "one-shot" transactions, they generate less trust than reciprocal exchanges. Thus, ironically, the very mechanisms that were created to reduce risk in transactions—the negotiation of terms and strictly binding agreements—have the unintended consequence of reducing trust in relationships.

BASIC CONCEPTS AND ASSUMPTIONS OF SOCIAL EXCHANGE

Social exchange occurs within structures of mutual dependence, in which actors are dependent on each other for valued outcomes. We assume that actors are motivated to obtain more of the outcomes that they value and others control, that actors provide each other with these valued benefits through exchange, and that exchanges between the same actors are recurring over time (rather than "one-shot" transactions). These three assumptions constitute the scope conditions for our analysis.

The simplest exchange relation consists of two actors, A and B, each of whom controls resources of value for the other. Questions of trust and commitment, however, are primarily of interest when A and B are embedded in a network of social relations that provides them with choices of alternative exchange partners. Our analysis focuses on *negatively connected* networks, in which each actor can choose among partners who are *alternative* sources of the same resource. In these networks, an actor's exchange with one partner is negatively correlated with her exchange with another.

Emerson's (1972a, 1972b) theory of power-dependence relations proposes that each actor's dependence is a source of power for the other; thus, A's power over B increases with B's dependence on A, and vice versa. B's dependence on A increases with the value of the outcomes that

A controls for B and decreases with the availability of those outcomes from B's alternative exchange partners. Inequalities in power and dependence create *power imbalanced* relations, in which the less dependent actor has a *power advantage* over the other that tends to produce a corresponding inequality in exchange benefits. We examine the development of trust within power imbalanced relations, comparing networks that vary substantially in the amount of imbalance they create and that provide opportunities for both benevolent and exploitative behavior.

Forms of Exchange

The mutual or reciprocal dependence in relations of exchange can be either direct (A provides value to B, and B to A) or indirect (the recipient of the benefit does not return benefit directly to the giver, but to another actor in the social circle). Our focus here is on direct exchange relations and the distinction that Blau (1964), Emerson (1981), and Lévi-Strauss (1969) all made between forms of direct exchange that are *reciprocal* or *negotiated*.

In *negotiated exchanges*, actors engage in a joint decision process, such as explicit bargaining, in which they reach an agreement on the terms of the exchange. Both sides of the exchange are agreed upon at the same time, and the benefits for both exchange partners are easily identified as paired contributions that form a discrete transaction. The flow of benefits in these exchanges is bilateral; that is, neither actor can obtain benefit without making an agreement that benefits both (however unequally). Most economic exchanges other than fixed-price trades fit in this category, as well as some social exchanges (e.g., agreements about the division of household labor). Research studying negotiated exchanges includes the work of Cook and associates (Cook and Emerson 1978; Cook et al. 1983; Friedkin (1993); Lawler and associates (Lawler 1992; Lawler and Yoon 1996); and Willer, Markovsky, and Skvoretz (Markovsky et al. 1993; Skvoretz and Willer 1993). In all of these programs, the agreements are also strictly binding; that is, the distribution of outcomes automatically follows an agreement.

In *reciprocal exchanges*, actors' contributions to the exchange are separately performed and nonnegotiated.² Actors initiate exchanges individually, by performing a beneficial act for another (such as assistance or ad-

² Fixed-priced trades, in which a price for a commodity or service is either explicitly set or implicitly assumed by convention, are also "nonnegotiated" but otherwise have the general characteristics of negotiated exchanges (bilateral flow of benefits and known terms). They have been of little interest to social exchange theorists and are excluded from our analysis.

vice), without knowing whether, when, or to what extent the other will reciprocate in the future. Because choices are made individually, benefits can flow unilaterally. Exchange relations develop gradually—or fail to develop—as beneficial acts prompt reciprocal benefit. Because the same act can complete one exchange and initiate another, discrete transactions are difficult to identify. Instead, the relation takes the form of a series of sequentially contingent acts; for example, your neighbor cares for your house while you are gone, you invite him to dinner, he gives you advice on buying a car, and so forth. The equality or inequality of these relations is established only over time, by the ratio of actors' individual giving to one another. Reciprocal exchanges are the focus of Molm's research program (e.g., Molm 1990, 1997) as well as much of the research on mixed-motive games; for example, Axelrod's (1984) work on the evolution of cooperation also studies how social interaction develops when actors make choices individually, without knowing each other's intentions and without negotiation.

Although negotiation is more typical of exchange in some settings (e.g., work) than in others (e.g., families), both forms of exchange are observed in a wide range of social contexts. Even in politics, business, and international relations, unilateral initiatives are common, and the expectation of future reciprocity is often left implicit. Similarly, even in interactions among family and friends, some exchanges of favors, household work, and choices of activities are negotiated. Thus, analytically, it is possible to distinguish the *form* of exchange from the actors, resources, or context of exchange. This is the strategy we take in our analysis.

The distinction between negotiated and reciprocal exchange is closely related to game theorists' distinctions between cooperative and noncooperative games. In cooperative games (and negotiated exchanges), strictly binding agreements are made jointly by actors who can communicate; in noncooperative games (and reciprocal exchanges), actors make choices individually, without knowledge of others' choices (Heckathorn 1985). Both types of games have been studied extensively by economists and rational choice theorists (e.g., Budescu, Erev, and Zwick 1999; Kagel and Roth 1995.).

Uncertainty and Risk

All forms of social exchange involve uncertainty and risk, but the amount and kind of risk vary. In reciprocal exchanges, actors initiate exchange without knowing what they are getting in return, and with no guarantee of the other's reciprocity. They may be able to infer the other's intentions once the relation is established, but their initial exchanges must take place

without that knowledge.³ Even in established relations, exploitation is always possible.⁴ Consequently, actors in reciprocal exchange relations risk giving benefits unilaterally while receiving little or no return. This is the type of risk—the risk of incurring a net loss—that is most critical to the development of trust, because it provides the opportunity for exchange partners to demonstrate their trustworthiness.

In negotiated exchanges, the bargaining process itself is a source of uncertainty; actors' choices of how hard to bargain, what tactics to use, and so forth all affect the terms of agreements and the likelihood of reaching an agreement. But once actors agree on the terms of an exchange, much of this uncertainty is eliminated: actors know what they are getting for what they are giving, and they can choose to engage in the exchange or not. The terms may be unequal and unsatisfactory to one or both parties, but unless both benefit more from the exchange than they would without it, it should not take place. If, in addition, the exchange is secured with conditions that make the agreement binding—conditions assumed in experimental research on negotiated exchange and characteristic of many negotiated exchanges in natural settings—the actors face no risk that the other party will not honor the terms of the agreement.

In natural settings, other sources of risk may make negotiated exchanges less secure. Agreements may not be strictly binding (Heckathorn 1985), time lags between promise and delivery may create opportunities for defection (Coleman 1990), and resource values may be uncertain (Kollock 1994). When these conditions hold, negotiated exchanges are also risky. In our analysis, we assume conditions that eliminate these sources of risk and that equate reciprocal and negotiated exchanges on these dimensions. First, we assume that exchange actions and agreements produce immediate benefits for actors in both forms of exchange (thus, in negotiated exchanges, agreements are binding); second, we assume that the value of benefits received from exchange is known and certain, and that the potential values of benefits from both forms of exchange are equal. These assumptions allow us to compare the development of trust in negotiated and reciprocal exchanges under the conditions that most contemporary exchange researchers have studied, and that represent most clearly the contrast in risk that the classical exchange theorists posed.

³ For this reason, Blau (1964) argued that the initial offer of a favor has special significance, because the person who makes the first overture risks not only the chance that the act will not be reciprocated, but rejection of the offer itself, i.e., the offer to enter into a relationship.

⁴ An excellent illustration of this point is the story that Hardin (1991, p. 185) relates from *The Brothers Karamazov*.

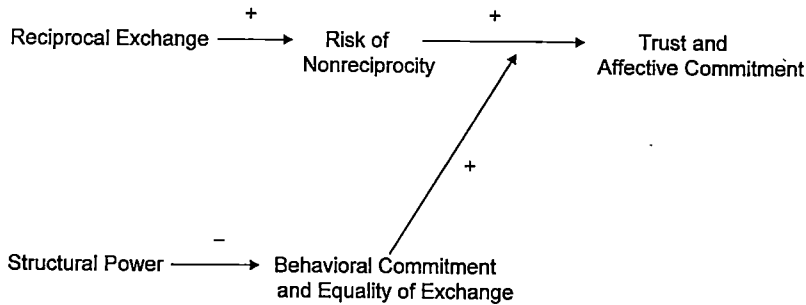


FIG. 1.—The theoretical model

TRUST, COMMITMENT, AND THE FORM OF EXCHANGE

We conceptualize trust and commitment as emergent phenomena that arise in response to uncertainty and risk. Blau (1964) proposed that they evolve together, beginning with minor transactions in which little trust is required because little risk is involved, and gradually expanding as partners prove themselves trustworthy and exchange with each other more frequently. Like Blau, we see trust and commitment as closely linked, but we distinguish between affective commitment and behavioral commitment and conceptualize their relations to trust somewhat differently.

Figure 1 illustrates the theoretical logic that we develop below. As the upper set of arrows shows, reciprocal exchange increases the risk of nonreciprocity that provides the opportunity for both trust and related feelings of affective commitment to develop. Whether trust actually develops, however, depends on behavioral indicators of actors' trustworthiness: their frequencies of exchange with each other rather than with alternative partners (behavioral commitment) and the equality of their exchange. The lower set of arrows proposes that these behaviors are influenced by the structure of power. We develop these arguments below.

Risk, Trust, and the Form of Exchange

We define *trust* as expectations that an exchange partner will behave benignly, based on the attribution of positive dispositions and intentions to the partner in a situation of uncertainty and risk. The conditions that promote trust occupy the middle regions between completely correspondent and completely noncorrespondent interests (Hardin 1991; Kelley and Thibaut 1978). If actors' interests are extremely noncorrespondent, there is little basis for the development of trust (or exchange), but if interests

are completely correspondent, there is no risk and no basis for inferring trustworthiness. Acts of trust, and attributions of trustworthiness, can be made only in situations in which the partner has both the incentive and the opportunity to exploit the actor, but instead behaves benignly. If, under these conditions, A behaves in a *trustworthy* manner, then B's *trust* in A should increase. A's untrustworthy behavior, on the other hand, should lead to B's distrust of A.

In one of the few empirical tests of this proposition, Kollock (1994) examined whether uncertainty in negotiated exchanges, produced by allowing sellers to deceive buyers about the quality of goods they were buying, promotes the development of trust. He found, as predicted, that ratings of partners' trustworthiness were greater when subjects negotiated exchanges under conditions of uncertain rather than certain quality. Disparities in trust were also greater, presumably because variations in trustworthy behavior affected the development of trust.

We use this same logic to test the prediction of the classical exchange theorists: that trust is more likely to develop in reciprocal exchanges than in negotiated exchanges. In developing this proposition, we draw on Yamagishi and Yamagishi's (1994) distinction between "assurance" and "trust." Unlike trust, which refers to expectations based on inferences about a partner's personal traits and intentions, *assurance* refers to expectations of benign behavior from an exchange partner based on knowledge of an incentive structure that encourages such behavior rather than exploitation.⁵ Mechanisms that provide assurance include legal or normative authorities that impose sanctions for violations of agreements or failure to fulfill one's obligations, guarantees such as collateral that protect against loss, warranties that assure certain standards of quality, and so forth. As long as an "assurance" structure is present, there is little opportunity for trust to develop, because there is little opportunity to learn about the partner and the partner's own dispositions and intentions.

Yamagishi and his associates (e.g., Yamagishi and Yamagishi 1994) use this distinction to explain why Japanese are less trusting of others than Americans (unlike American society, Japanese society has extensive assurance structures that guarantee mutual cooperation) and why Japanese subjects in social dilemma experiments are less likely to stay in the experimental group when given the opportunity to exit (in the absence of the typical Japanese assurance structures, Japanese subjects are less likely to trust that group members will not free ride). Similarly, we propose that trust is more likely to develop in reciprocal exchanges than in negotiated

⁵ Others have made similar distinctions. For example, Hardin (1991) distinguishes between "trust" and "reliance," and Dasgupta (1988) between "trust" and "confidence."

exchanges, because negotiated exchanges provide "assurance," while reciprocal exchanges enable "trust."⁶

In negotiated exchanges, joint decision making informs actors of the benefits they will receive from the exchange, and agreements, when binding, guarantee that those benefits are delivered. Although uncertainty in the bargaining process itself remains, that form of uncertainty should have less bearing on the development of trust. If actors in negotiated exchanges offer too much and strike a poor bargain, or lose out to a competitor who has offered more, they are more likely to make inferences about the bargaining skills of the actors than about the trustworthiness of the partner. Even when actors make agreements that are less favorable to themselves than they had hoped, they do so knowingly, and trust is not an issue to the extent that agreements are binding.

In reciprocal exchanges, the only form of assurance comes from the expectation of future interaction. When exchanges between the same actors are recurring over time, as our analysis assumes, then the "shadow of the future"—the knowledge that the other can respond contingently and in kind to one's own behavior—provides some protection (Axelrod 1984; Schelling 1960).⁷ As long as actors value the continuation of a relation (or fear retaliation), they are less likely to exploit one another. This form of assurance, however, is weaker than binding agreements with known terms, and it develops only over time.

We propose that the weaker assurance and greater risk of exploitation in reciprocal exchange provides greater opportunity for actors to demonstrate trustworthiness to one another. To the extent that they do, then levels of trust in reciprocal exchanges should be greater, on average, than in negotiated exchanges. But the development of trust also depends on how the partner responds to the opportunity to exploit or reciprocate. Trust should increase with the partner's reciprocity and, more generally, with the development of frequent, stable exchange between the two actors (Dasgupta 1988; Lawler and Yoon 1996). While these behaviors might be expected to increase trust in both forms of exchange, they should have a stronger effect in reciprocal exchanges. Actors in reciprocal exchanges are more likely than those in negotiated exchanges to attribute the partner's behaviors to personal traits of the partner (such as trustworthiness/untrustworthiness) rather than to the assurance structures that surround the exchange.

⁶ Our analysis focuses on the development of trust in specific exchange partners, however, whereas Yamagishi's analysis applies to the development of trust in general others.

⁷ When the same actors repeatedly engage in reciprocal exchanges with each other, they can employ conditional strategies such as "tit-for-tat" that reward the other's exchange and punish their defection, thus reducing risk and increasing assurance.

Because the “shadow of the future” provides some assurance, it should make trustworthy behaviors at least as likely as untrustworthy ones. Thus, under the conditions of recurring exchange assumed for our analysis, levels of trust should be higher in reciprocal than in negotiated exchanges. Under other conditions, however—such as a very short shadow of the future, combined with structural incentives that encourage exploitative behavior—*un*trustworthy behavior might be the norm, and trust might then be lower in reciprocal than in negotiated exchanges. Transient exchanges between strangers, for example, are particularly risky (Macy and Skvoretz 1998).

In our study, we examine actors’ trust in their exchange partners, and its relation to the partners’ trustworthy or untrustworthy behavior, under structural conditions that vary in the incentives they provide for benign or exploitative behavior. As long as future interaction between the same actors is expected, we predict that trust will be greater, on average, in reciprocal than in negotiated exchanges, although the levels of trust should vary with structure.

Commitment and Trust

Exchange theorists have conceptualized the development of commitments between partners in two ways: (1) as purely *behavioral* patterns of exchange in which pairs of actors choose to exchange repeatedly with one another rather than with alternative partners, primarily as a means of reducing uncertainty (Cook and Emerson 1978; Kollock 1994), and (2) as *affective* bonds that develop from repeated experiences with successful exchanges between the same partners (Lawler and Yoon 1996).

We propose that both forms of commitment are related to trust, but in different ways. As figure 1 illustrates, behavioral commitments are one form of “trustworthy” behavior that should encourage the development of trust in a context of uncertainty and risk. An actor who repeatedly exchanges with another, forgoing alternative partners, signals his or her trustworthiness as an exchange partner; one who defects repeatedly does the opposite. Repeated exchanges with the same partner also provide information about the other that reduces uncertainty and makes the other’s behavior more predictable. More generally, they are part of the process of “embeddedness” that Granovetter (1985) has addressed, in which ongoing social relations can contribute to the production of trust. As we discuss in more detail below, behavioral commitments can also reduce power use and inequality (Cook and Emerson 1978), which should provide further indication of the partner’s trustworthiness—particularly for powerful partners.

Whereas behavioral commitments contribute to the development of

trust, we propose that affective commitments emerge from the same processes that produce trust (fig. 1). We define affective commitments as feelings of liking for, and attachment to, a specific exchange partner, which are indicated by expressions of commitment to the partner and positive evaluations of the partner.⁸ Affective commitments may be more or less closely related to actual behavioral commitments. We propose that these feelings of affective attachment to the partner are generated through the same processes that produce trust. That is, in the absence of assurance structures that can explain a partner's positive behaviors, actors are more likely to attribute the behaviors to the partner's personal traits and intentions. If so, then positive feelings of affect and commitment toward the partner will, like trust, tend to be stronger in reciprocal than in negotiated exchanges. They will increase with the partner's behavioral commitment and with the equality of their exchanges, and these behaviors will have a stronger effect on feelings of affect and commitment in reciprocal than in negotiated exchanges.⁹

Power, Behavioral Commitments, and Inequality

At the same time that behavioral commitments reduce uncertainty and reduce transaction costs, they *increase* opportunity costs. Actors who are committed to a single partner obtain security and predictability at the cost of forgoing the opportunity to explore other, possibly better alternatives. For this reason, commitments are more likely to be used as a means of reducing uncertainty when alternatives are limited and opportunity costs are low (Yamagishi, Cook, and Watabe 1998). Consequently, as figure 1 proposes, behavioral commitments should vary inversely with power (Cook and Emerson 1978; Lawler and Yoon 1996; Tallman, Gray, and Leik 1991). Low-power actors (who have few or poor alternatives) should be more likely to make commitments than high-power actors (who have more or better alternatives), and high-power actors should be more likely to make commitments the weaker their power advantage.

To examine the effects of power on commitment and to produce systematic variations in both commitment and equality (our indicators of trust-

⁸ Our conceptualization differs slightly from Lawler and Yoon's (1996). They define affective commitment as an emotional or cathectic attachment to the relation or group, characterized by positive emotions of pleasure and satisfaction, or interest and excitement. In contrast, our focus is on attachment to a specific other, characterized by positive evaluations and feelings for the other.

⁹ Our linkage of trust with affective bonds is similar to Gambetta's (1988, p. 232) notion of "thick trust": trust that is reinforced by feelings of liking and affection. Lewis and Weigert (1985) argue that all types of trust have an affective component, although this component is most intense in close interpersonal trust.

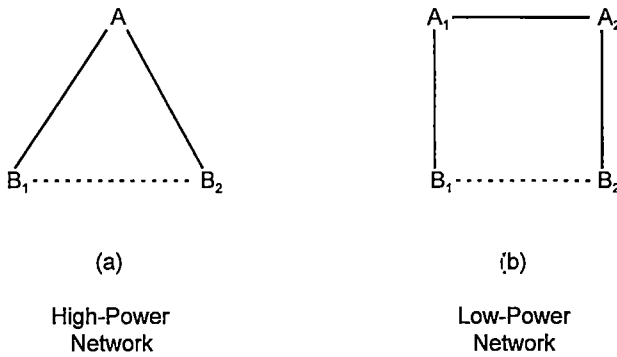


FIG. 2.—Networks varying A's power advantage over B. Following Cook and Emerson's (1978) convention, solid lines indicate potential relations with high exchange value, and dotted lines indicate potential relations with low exchange value.

worthy behavior), we compare the development of trust among actors whose exchange relation is embedded in one of the two simple networks shown in figure 2. In both networks, actors occupy one of two structural positions, A and B. The exchange relation *between* A and B—the behavioral commitment, trust, and affective attachment that develop between A and B—is the focus of our study. This power-imbalanced relation offers “mixed motives” for interaction; that is, both A and B can benefit from exchange with one another, but their interests are not perfectly aligned. The extent to which they are aligned varies with the actors' positions (A and B) within each network and with the variations in structural power across the networks. These differences in power and position create variations in opportunities and incentives for trustworthy or untrustworthy behavior.

Within each network, we create a power advantage for A over B, by giving A a high-value alternative to exchange with B (the other B in fig. 2a, and the other A in fig. 2b), while B has only a low-value alternative to exchange with A (the other B in both networks). Across networks, we vary the strength of A's power advantage by varying the *availability* of A's alternatives (Cook et al. 1983). In figure 2b, the addition of a fourth actor to the network reduces the availability of A's alternative to B; that is, A's alternative to B in figure 2b (another A) is less available than A's alternative in figure 2a (another B), because the other A, unlike the B, has a high-value alternative and is less dependent on A. Consequently, A's power advantage is weaker in 2b than in 2a.¹⁰

¹⁰ Markovsky et al. (1993) refer to networks that differ in the availability of alternatives for powerful actors as “strong power” networks and “weak power” networks.

Because A has a more valuable alternative to B than B has to A in both networks, the formation of a commitment between A and B is more costly for A than for B. Consequently, although A's and B's commitments to each other are necessarily related, A's behavioral commitment to B should be lower than B's commitment to A in both networks. Reducing A's power advantage reduces the opportunity costs for A of making a commitment; consequently, commitments between A and B should be more likely to form in the low-power networks than in the high-power networks, and exchanges between A and B should be more equal.

The Form of Exchange, Behavioral Commitment, and Inequality

Finally, we consider whether the form of exchange affects behavioral commitments and inequality. We know from earlier work that reciprocal exchanges tend to be more equal than negotiated exchanges (Molm, Peterson, and Takahashi, in press). We might also expect that if behavioral commitments reduce uncertainty and risk, then actors in reciprocal exchanges—who face greater risk of exploitation—might be more likely to form committed relations than those engaged in negotiated exchanges. But two considerations make this prediction unlikely.

First, behavioral commitments reduce not only the uncertainty of reciprocity in reciprocal exchange, but the uncertainty of bargaining processes in negotiated exchanges. Actors who bargain repeatedly with each other become familiar with each other's tactics and come to terms more quickly and easily. Even more importantly, behavioral commitments eliminate the risk that actors will lose out in competition with others who are negotiating for an agreement with the same partner. A supplier who secures a regular contract with a buyer, for example, is in a more advantageous position than one who has to compete with other suppliers.

Second, behavioral commitments are more costly, in general, to actors in reciprocal exchanges than to actors in negotiated exchanges. Because actors in reciprocal exchanges make choices individually and benefits can flow unilaterally, it is possible for them to receive benefits from multiple partners at the same time. Consequently, these actors will maximize their benefits to the extent that they can maintain exchanges with multiple partners, by intermittently reciprocating their giving. Thus, in reciprocal exchanges, commitments are costly for powerful actors even in low-power networks like figure 2b.

For these reasons, we do not expect behavioral commitments to be more common in reciprocal than in negotiated exchanges. To the extent that the form of exchange affects either the frequency or inequality of exchange, however, these effects could contribute to a relation between exchange form and trust and must be controlled in our analysis.

Hypotheses

Our central predictions, shown in figure 1 and discussed above, can be summarized in the following hypotheses:

HYPOTHESIS 1.—*An actor's trust and affective commitment for an exchange partner will be greater, on average, in reciprocal exchange relations than in equivalent (structurally and behaviorally) negotiated exchange relations.*

HYPOTHESIS 2.—*An actor's trust and affective commitment for a partner will increase with the partner's behavioral commitment to the actor and decrease with the inequality of their exchange; these relations will be stronger in reciprocal than in negotiated exchanges.*

HYPOTHESIS 3.—*Actors' behavioral commitments to their partners and the equality of their exchange will vary inversely with their power.*

3a.—*Behavioral commitments will be greater, and inequality lower, when the power imbalance of a relation is low (fig. 2b) rather than high (fig. 2a).*

3b.—*Within relations, the behavioral commitment of the less powerful actor will be greater than that of the more powerful actor.*

METHOD

We test these hypotheses in a laboratory experiment designed to meet the traditional scope conditions of social exchange theory (Molm and Cook 1995) and our analysis: actors are mutually dependent on one another for valued outcomes (operationalized as money), they are recruited on the basis of their interest in acquiring more of those outcomes, and they engage in repeated exchanges with one another over time. All actors have alternative exchange partners, thus allowing variability in commitments—both behavioral and affective—to a given partner and offering the opportunity for both benign and exploitative behavior.

Experimental Design and Basic Procedures

A 2×2 factorial design crossed the form of exchange (reciprocal or negotiated) with the power advantage for the powerful actor (high power or low power). In each of the four resulting conditions, 10 networks were run, with 140 undergraduate students randomly assigned to conditions and to positions within networks. All networks consisted of same-sex subjects, with gender balanced within experimental conditions for control purposes.¹¹

¹¹ Because preliminary data analyses found no effects of gender on the measures examined here, gender is omitted from these analyses.

Subjects participated in one of the two networks shown in figure 2, and they engaged in either reciprocal or negotiated exchanges to earn money.¹² To provide the opportunity for commitment and trust to develop, subjects remained in the same network positions throughout the experiment and interacted repeatedly with the same two partners. All interaction occurred through computers; subjects were seated in isolated rooms and never met each other. To provide subjects with alternative partners, relations in the networks were negatively connected, and these connections were operationalized in the traditional way (Cook and Emerson 1978): in negotiated exchanges, an agreement with one exchange partner precluded an agreement with another on that opportunity, and in reciprocal exchanges, initiating exchange with one actor precluded initiating exchange with another on that opportunity.¹³

Following detailed instructions and practice trials, subjects participated in a series of exchange opportunities. At the end of each opportunity, they were informed about the source and amount of any points gained, and their total earnings (points \times cents) were cumulated and shown on their computer screens.

Subjects were not informed in any of the conditions about the amounts of money their partners received from exchanges with them or others, the outcomes of their partners' exchanges with other actors in the networks, or their partners' cumulative earnings. These restrictions on subjects' information, along with other restrictions discussed below, served several purposes: they reduced competition and equity as motives, they enhanced the uncertainty in which all exchanges—both negotiated and reciprocal—were conducted, and they created conditions more similar to those of exchanges in natural settings.¹⁴

¹² Money is used as the valued outcome in our experiment (and in virtually all exchange experiments) solely because of its advantages for experimental control: money is widely valued, it can be quantified to produce a ratio level of measurement, and it is resistant to the effects of satiation or diminishing marginal utility (which would alter value). The exchange resource in the experiment is not money per se—money is not transferred from one actor to another, as in economic exchanges—but rather the capacity to produce valued outcomes, operationalized as money, for another.

¹³ The ability of actors in reciprocal exchanges to *receive* rewards from more than one partner at the same time does not negate the negative connection. Because rewards received are always some function of rewards given—even when reciprocity is intermittent—increased exchange in one relation still reduces exchange in another, thus satisfying the definition of negatively connected relations.

¹⁴ Power-dependence researchers (e.g., Cook and Emerson 1978; Lawler and Yoon 1996) commonly reduce subjects' information about their partners' outcomes to minimize their concerns with either equity or competition, and to focus their attention on their own outcomes. In natural settings, actors rarely know exactly how much benefit each is receiving from an exchange, and so precise comparisons with others' outcomes are impossible. This is true even in most economic exchanges. For example, when

Manipulations

The form of exchange.—Two different exchange settings were created to manipulate the form of exchange, negotiated or reciprocal. The settings were designed to be as comparable as possible to one another on all dimensions other than their defining differences.

The negotiated setting shares many of the features of other negotiated exchange settings but is most similar to Lawler and Yoon's (1998) setting. Actors negotiated the division of a fixed amount of benefit within relations on each of a series of exchange opportunities, and each opportunity consisted of up to five rounds of negotiation. On each round, all actors in the network simultaneously made offers to all alternative partners. After the first round, actors could accept another's offer, repeat their last offer, or make a counteroffer. Negotiations continued until all potential agreements were made or the five rounds were up. As soon as an agreement was reached, both actors received the amounts they had agreed upon (thus, agreements were binding).

Subjects knew the range of points they could request from agreements and that, in general, the more they received, the less the other person received. They did not know that a fixed amount of profit was divided, however, nor did they know how many points the other subject received from an agreement. Subjects made offers by *requesting* the number of points they wanted to receive from an agreement, and each subject's *request* for points was then converted, by the computer, into an *offer* of the remaining points for the other subject. Subjects were told that they would not know how many points the other person received and that they should be concerned only with their own points.

The reciprocal exchange setting is similar but not identical to the setting used by Molm in previous work (e.g., Molm 1990, 1997). Each actor in the network gave points to one of his or her exchange partners on each of a series of exchange opportunities. To hold constant the potential joint benefit of reciprocal and negotiated exchanges, the number of points that each actor could give to any partner on any single opportunity was fixed and equal to one-half the total points that actors in the negotiated exchange setting could divide on each opportunity. If, for example, subjects in a negotiated exchange relation could divide 16 points on an exchange opportunity, subjects in an equivalent reciprocal exchange relation could

purchasing a car from a dealer, most buyers do not know exactly how much profit the dealer receives, and the dealer does not know exactly how much the buyer values the car. What actors do know, typically, is that the more they offer or give, the more the other receives and the less profit they receive. And that is what our subjects knew, as well.

give each other 8 points on an opportunity.¹⁵ Inequality in these relations thus resulted from unequal reciprocities over time, rather than from unequal benefits within agreements.

As in the negotiated exchange conditions, subjects knew only the number of points they could receive from others, not the number of points they could give to others. On each opportunity, all subjects simultaneously and independently chose a partner to give points to, without knowing how many points they gave and without knowing whether or when the other would reciprocate. Subjects were then informed that each of their partners either added to the subject's earnings or did not act toward the subject, and their cumulative points were updated.

Subjects in the negotiated exchange conditions exchanged for 100 opportunities (the maximum that could be completed in a two-hour session); those in the reciprocal exchange conditions, for 400 opportunities. Both sessions took roughly the same amount of time, and subjects were not informed of the number of opportunities in either. More opportunities were run in the reciprocal exchange conditions to allow for the possibility that reciprocal exchange takes longer to stabilize than negotiated exchange, which consists of multiple behaviors—up to five rounds of offers—and more information on each opportunity. The monetary value of points was adjusted so that subjects in the reciprocal exchanges earned the same amount as subjects in the negotiated exchanges for comparable behaviors.

The structure of power.—We manipulated power by comparing the two networks shown in figure 2. In both networks, A's more valuable alternative to B gives A a power advantage over B. The strength of A's advantage is greater in the high-power network (2a) than in the low-power network (2b), and therefore the opportunity costs to A of a behavioral commitment to B are higher in 2a than in 2b. All other variables affecting power are held constant; in both networks, all actors have two alternative partners, the potential value of the A–B relation is high and constant, and the potential values of A's and B's alternative partners are the same.

"High-value" relations in the networks were worth 16 points, and "low-value" relations, 4 points. Actors in the negotiated exchange conditions negotiated the division of 16 or 4 points, while actors in the reciprocal exchange conditions could give 8 points to partners in high-value relations and 2 points to those in low-value relations.

Because our interest was in the effects of the *behaviors* produced by

¹⁵ This procedure assured that the expected values of exchange benefits were equal in the negotiated and reciprocal conditions under baseline conditions (equal division of resources and equal probability of agreements when exchange is negotiated, and equal probabilities of giving by all actors to all partners when exchange is reciprocal).

these structures on attributions of trust and evaluations of the partner, we eliminated the structure itself as a source of attributions.¹⁶ In all conditions, subjects' information about the structure of the networks was restricted to knowledge of their own access to their two exchange partners and their potential benefits from these partners. Because this information was identical for both networks, subjects' perceptions of the network structures should be identical (at least initially).

Measures

Subjects' evaluations of the partner: trust and affective commitment.—At the conclusion of the experiment, subjects responded to a series of seven-point bipolar semantic differential scales, measuring evaluations of their exchange partners and relations. We derived three measures of trust and affective commitment from these items, using A's evaluations of B and B's evaluations of A.

We measured trust with a single item asking subjects, "How much did you trust Person X during the experiment (very much/very little)?" We assessed affective commitment with two measures, one a single item asking subjects how committed/uncommitted they felt toward the partner, and the other a five-item scale measuring their affective evaluation of the partner. The first four items of the scale asked subjects to evaluate their partner's behavior as good/bad, nice/awful, cooperative/uncooperative, and helpful/unhelpful; the fifth item asked subjects to describe their general feelings toward the partner as positive/negative. With responses to the five items averaged, the resulting scale has alpha reliabilities of .97 and .95 for A and B, respectively. All three measures range in value from 1 to 7, with higher values indicating stronger feelings of trust and commitment and more positive affect.¹⁷

Partner's behavior toward the subject: behavioral commitment and inequality.—We measure the "trustworthiness" of the partner's behavior with two indicators: the partner's behavioral commitment to the subject and the equality or inequality of exchanges between the partner and the subject.

¹⁶ When actors know the relative power positions that they and others occupy, that knowledge alone might influence their expectations of the partners' trustworthiness or interpretations of the partners' behavior. Because subjects in our experiment were unaware of the power positions of actors in their network, such expectations and inferences were impossible.

¹⁷ Correlations among the three variables range from a high of .87 for the relation between A's feelings of commitment to B and A's positive affect toward B, to a low of .53 for the relation between B's trust in A and B's feelings of commitment to A. All correlations are higher, without exception, for the reciprocal exchanges than for the negotiated exchanges.

We used a modified version of Cook and Emerson's (1978) measure to examine the partner's (P) behavioral commitment to the subject (S); this measure computes how often P chose to exchange with S rather than with an alternative partner. For negotiated exchanges, P's commitment to S equals:¹⁸

$$C_P = [\Sigma(T_{PS} - T_{Palt})]/T_P,$$

where T_{PS} = transactions (agreements) between P and S; T_{Palt} = transactions between P and P's alternative partner; and T_P = the total number of transactions that P completes with both S and P's alternative partner. For reciprocal exchanges, T_{PS} and T_{Palt} are replaced by the frequency of P's giving to S (G_{PS}) and to P's alternative partner (G_{Palt}), and T_P is replaced by the frequency of P's giving to both partners, which equals the number of exchange opportunities. The commitment measures have a maximum value of +1 when all of P's transactions or givings are with S, a minimum value of -1 when all of P's transactions or givings are with the alternative partner, and a value of 0 when P's transactions or givings are equally distributed between both partners.

We also measured the inequality (IE) of the value of the exchanges between A and B:

$$IE_{AB} = (V_{AB} - V_{BA})/(V_{AB} + V_{BA}),$$

where V_{AB} = the total points A received from B, and V_{BA} = the total points B received from A. The measure has a potential range of -1 (maximum inequality in B's favor) to +1 (maximum inequality in A's favor), with 0 indicating equal exchange.¹⁹ Although subjects were not informed of the absolute inequality of their exchanges, those in reciprocal exchanges knew whether their giving was reciprocated, and those in negotiated exchanges knew that the more points they received from an agreement, the fewer points their partner received.

Analytic Strategy

Our analysis takes account of the dependencies in our data (between actors within networks, and between behaviors across exchange opportunities) in two ways. First, the statistical unit for all of our analyses is the exchange network; thus, all analyses are based on an N of 40. Second, our measures of behavioral commitment and inequality consist of overall

¹⁸ For subjects in the A position, P refers to B; for subjects in the B position, P refers to A.

¹⁹ Although the potential range of our inequality measure is -1 to +1, its practical range is 0 (equality) to +1 (inequality in favor of the power-advantaged actor). Inequality favored the power-disadvantaged actors in only a few cases (10%), and their advantage was small.

TABLE 1
MEANS AND STANDARD DEVIATIONS OF SUBJECTS' TRUST AND AFFECTIVE
COMMITMENT TOWARD THE PARTNER, BY EXPERIMENTAL CONDITION

MEASURES	RECIPROCAL EXCHANGE		NEGOTIATED EXCHANGE	
	High-power Network	Low-power Network	High-power Network	Low-power Network
Trust in partner:				
A's trust in B	4.20 (1.48)	6.00 (1.68)	3.80 (1.48)	4.25 (1.44)
B's trust in A	3.80 (1.62)	5.90 (.77)	3.40 (1.78)	3.65 (.91)
Positive affect for partner:				
A's affect for B	4.78 (1.44)	6.17 (1.37)	4.34 (1.25)	4.64 (1.46)
B's affect for A	4.36 (1.43)	6.02 (1.06)	3.10 (1.16)	3.75 (.57)
Feelings of commitment to partner:				
A's commitment to B	4.80 (1.69)	6.00 (1.81)	4.00 (1.41)	5.20 (1.90)
B's commitment to A	5.30 (1.49)	6.00 (1.13)	3.40 (1.90)	4.80 (1.58)

NOTE.—Higher values indicate greater trust, more positive affect, and stronger feelings of commitment, on a scale of 1 to 7. SDs are given in parentheses.

frequencies or values, computed for the entire exchange period or (when specified) for the last half of the exchange period. In the low-power networks, measures of all evaluative and behavioral variables were computed for each A and each B and then averaged for the two actors in each position. In the high-power networks, A's relation with one of the two Bs was randomly selected for the analysis, and measures were computed for that relation.²⁰ Mean values of these measures, computed for the entire exchange period, are reported in tables 1 and 2.

RESULTS

Analyses of Behavioral Commitment and Inequality

We first examine whether our manipulation of power created the effects on behavioral commitment and inequality predicted in hypothesis 3. As

²⁰ In the high-power networks, in which a single A chooses between exchange with the two Bs, A's behavior toward one B is perfectly and negatively correlated with A's behavior toward the other B. Therefore, averaging A's behavior toward the two Bs, or the Bs' evaluations of A, would be misleading.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF SUBJECTS' BEHAVIOR TOWARD PARTNER, BY
EXPERIMENTAL CONDITION

BEHAVIOR	RECIPROCAL EXCHANGE		NEGOTIATED EXCHANGE	
	High-power Network	Low-power Network	High-power Network	Low-power Network
A's behavioral commitment to B05 (.32)	.32 (.32)	-.07 (.34)	.61 (.37)
B's behavioral commitment to A32 (.29)	.43 (.27)	.30 (.47)	.65 (.35)
Inequality in the A-B relation12 (.08)	.05 (.05)	.39 (.33)	.15 (.20)

NOTE.—SDs are given in parentheses.

table 2 shows, the mean values for behavioral commitment varied considerably, from a low of $-.07$ for A in the high-power negotiated condition, to a high of $.65$ for B in the low-power negotiated condition. A multivariate analysis of variance on these means, using power and form of exchange as between-subjects variables and actor (A or B) as a within-subject variable, found significant main effects on commitment of both power and actor, as well as a significant interaction between the two ($F[1,36] = 11.35$ for power, 37.83 for actor, and 15.24 for power \times actor, all significant at $P < .01$). As hypothesis 3a predicted, behavioral commitments between A and B were more frequent in the low- than in the high-power network; not surprisingly, this effect was stronger for A than for B. And as hypothesis 3b predicted, B's commitment to A was greater than A's commitment to B in all conditions (see table 2).

As expected, the form of exchange had no significant main or interactive effects on behavioral commitment. These commitments tended to be somewhat weaker in the low-power networks when exchange was reciprocal rather than negotiated, but the difference was not significant. As figure 3 shows, however, this condition—the low-power network with reciprocal exchange—differed markedly from the other three in the *change* in commitment that occurred over the course of the exchange period. The average commitment of A and B to each other increased significantly over time in this condition ($t = 2.34$; $P < .05$, two-tailed test), but not in any of the others. We return to this difference in our analysis of trust and affective commitment.

The inequality of exchange in the A-B relation is significantly influenced by both power and form of exchange, in the directions expected ($F[1,36] = 6.26$, $P < .05$ for power; $F[1,36] = 8.74$, $P < .01$ for form).

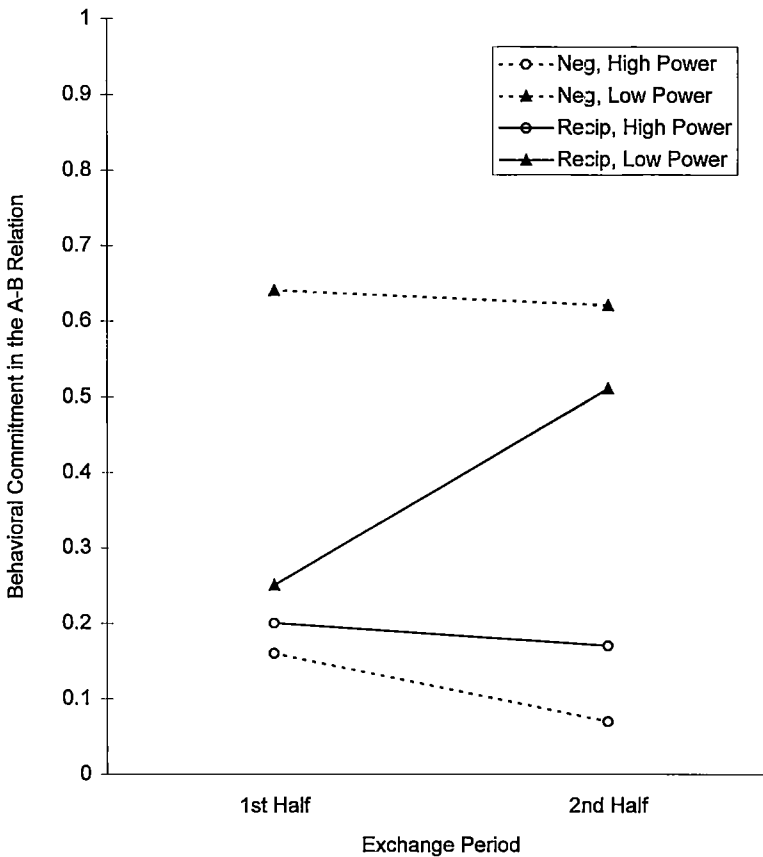


FIG. 3.—Behavioral commitment in the A-B relation

Inequality is greater when A's power advantage is greater, as virtually all theories of power in exchange networks predict, and greater when exchange is negotiated rather than reciprocal, as Molm et al. (in press) found previously. The average negotiated agreement between A and B in the high-power networks gave A twice as many points as B (11 points for A and 5 for B); in the low-power networks, both reciprocal and negotiated exchanges were more equal. Unlike behavioral commitment, inequality did not change significantly over time in any of the conditions.

The relationship *between* behavioral commitment and inequality varies substantially by the form of exchange. Theoretically, commitment should reduce the use of power and inequality in both forms of exchange; empirically, a negative relation between commitment and power use holds only

for reciprocal exchanges.²¹ Behavioral commitment and inequality are negatively correlated when exchange is reciprocal ($r = -.73$ for A and $-.36$ for B) but unrelated when exchange is negotiated ($r = .08$ for A and $-.06$ for B). Because only one of these correlations is significant, we examine the effects of both variables in our remaining analyses. For these analyses, we use measures of behavioral commitment and inequality during the last half of the exchange period, the interval that is most likely to affect measures of subjects' trust and affective commitment at the conclusion of the exchange.

Analyses of Trust and Affective Commitment

Our central hypotheses concern the effects of the form of exchange and the partner's behavior on the development of trust and affective ties to the partner. We predict that trust in the partner and related feelings of affect and commitment will be greater, on average, in relations of reciprocal exchange than in relations of negotiated exchange (hypothesis 1), and that the partner's behavior (their behavioral commitment and the inequality of exchange) will have stronger effects on trust and affective commitment in reciprocal exchanges than in negotiated exchanges (hypothesis 2).

Table 3 reports the results of analyses of variance on each of the three dependent measures (trust, positive affect, and feelings of commitment) for both the more (A) and less (B) powerful actors. For each measure, we first present the results of a standard analysis of variance on our factorial design, and then present the results of an analysis of covariance in which the two behavioral measures—behavioral commitment to the partner and the inequality in the relation—are entered as covariates before the effects of the experimental variables are assessed. The analysis of covariance allows us to test whether the form of exchange still affects trust and affective commitment after controlling for any differences in behavioral commitment and inequality between the two forms of exchange, as hypothesis 1 predicts.

The results in table 3 strongly support hypothesis 1: trust in the partner and related feelings of positive affect and commitment are greater in reciprocal exchange relations than in negotiated exchange relations, for both A and B. As table 1 shows, this is true for every comparison of means for the two forms of exchange (only the effect of the form of exchange on A's

²¹ Although exchange theorists typically assume that commitment drives this relation, our experimental design does not allow us to determine the direction of causality. Both behavioral commitment and inequality are measured variables that develop throughout the exchange period, and it is quite possible that actors who use power less are more likely to form committed relationships.

TABLE 3

ANALYSES OF VARIANCE AND COVARIANCE ON SUBJECT'S FEELINGS OF TRUST, AFFECT, AND COMMITMENT TOWARD THE PARTNER

SOURCE	TRUST IN PARTNER				POSITIVE AFFECT TOWARD PARTNER				FEELINGS OF COMMITMENT TO PARTNER			
	A's Trust in B		B's Trust in A		A's Affect for B		B's Affect for A		A's Commitment to B		B's Commitment to A	
	ANOVA F(1,36)	ANCOVA F(1,34)	ANOVA F(1,36)	ANCOVA F(1,34)	ANOVA F(1,36)	ANCOVA F(1,34)	ANOVA F(1,36)	ANCOVA F(1,34)	ANOVA F(1,36)	ANCOVA F(1,34)	ANOVA F(1,36)	ANCOVA F(1,34)
Covariates:												
Commitment	18.25***	..	2.90*	...	23.82***	.	12.40***	...	27.39***	.	12.92***
Inequality	2.28*	...	2.75*	..	.22	...	9.07**	...	3.41*12
Main Effects:												
Form of exchange (F)	4.99*	6.23**	9.74***	8.14**	5.08*	9.31**	25.74***	23.28***	2.18	2.43	9.99**	12.57***
Structural power (P) ..	5.47*	1.08	7.66**	3.60*	3.74*	.66	11.02***	1.90	4.90*	28	4.58*	.40
Interaction Effect:												
F × P	1.97	3.96*	4.74*	5.43**	1.56	2.95*	2.11	5.01*	.00	.27	.51	.23

NOTE.—The analyses of covariance are conducted using the classical experimental approach; i.e., all covariates are entered first, then all main effects, then the interaction term.

* $P < .05$.** $P < .01$.*** $P < .001$, one-tailed tests.

TABLE 4

CORRELATION COEFFICIENTS BETWEEN PARTNER'S BEHAVIORS AND SUBJECT'S
FEELINGS TOWARD THE PARTNER, BY FORM OF EXCHANGE

PARTNER'S BEHAVIOR	A'S FEELINGS TOWARD B			B'S FEELINGS TOWARD A		
	Trust	Affect	Commit	Trust	Affect	Commit
Behavioral commitment: ^a						
Negotiated exchange48*	.44*	.55**	-.07	.27	.51*
Reciprocal exchange75***	.89***	.86***	.60**	.71**	.55**
B for interaction ^b	2.21*	2.51**	2.23*	2.16**	1.63**	-.25
Inequality between A and B:						
Negotiated exchange15	.33	.01	-.01	-.02	.17
Reciprocal exchange	-.80***	-.74***	-.91***	-.43*	-.65***	-.31
B for interaction ^b	11.39***	9.94***	12.35***	5.20*	7.18***	4.13

^a B's commitment to A is correlated with A's feelings for B, and A's commitment to B is correlated with B's feelings for A.

^b Unstandardized regression coefficients for the interaction between form of exchange and behavior. The multiplicative interaction terms were entered along with their component terms in each equation.

* $P < .05$.

** $P < .01$.

*** $P < .001$, one-tailed tests.

feelings of commitment to B fails to reach significance in table 3; $P = .06$). When exchanges are reciprocal, all but one of the mean ratings of subjects' trust, affect, and commitment for the partner are above the neutral point of 4.0; when exchanges are negotiated, half of the means are below 4.0. Even in the condition that produces the lowest levels of trust—B's trust in A in the high-power networks—the mean value for reciprocal exchange, while below the neutral point of 4.0 (3.8), is still higher than the comparable mean for negotiated exchange (3.4).

As comparisons of the two sets of analyses in table 3 show, the effects of the form of exchange are still significant—and, in some cases, stronger—after controlling for the effects of the partner's behavioral commitment and the inequality of exchanges in the relation. In contrast, the effects of structural power are greatly reduced, as we would expect.

Hypothesis 2 predicts that trust and affective commitment increase with the partner's behavioral commitment and decrease with the inequality of their exchange, with these relations stronger in reciprocal than in negotiated exchanges. The results of the covariance analysis in table 3 support the predicted effects of these behaviors (particularly behavioral commitment) on the development of trust and affective ties; table 4 examines whether their effects interact with the form of exchange. Because of the significant correlation between the two behavioral variables in reciprocal exchanges, we examine the effects of each behavior separately, re-

porting both their zero-order correlations with trust and affective commitment, and the regression coefficients for the interaction between each behavior and the form of exchange.

The results strongly support hypothesis 2: all of the correlations between the partner's behaviors and the subject's feelings of trust, affect, and commitment for the partner are stronger for reciprocal exchanges than for negotiated exchanges. Regression coefficients for the interactions of behavior with form of exchange show that all but two of these differences are significant. In reciprocal exchanges, the development of trust and affective ties are strongly influenced by the partner's behavior, increasing with the partner's behavioral commitment and decreasing with the inequality of their exchange. In negotiated exchanges, inequality has no significant effects on trust or affective commitment, and the effects of behavioral commitment are weaker and primarily affect A's feelings toward B.

The interaction of power with the form of exchange.—Finally, we turn to an unpredicted but intriguing finding: the interaction between power and the form of exchange on feelings of trust and positive affect (table 3). Power and the form of exchange interact significantly in their effect on B's trust in A in the original analysis of variance and in their effects on both actors' trust and affect for the other once the behavioral covariates are controlled. The means in table 1 illustrate the pattern that produces these effects: power has a stronger effect on both actors' trust and affect when exchange is reciprocal than when it is negotiated; in particular, the means in the low-power networks are much higher when exchange is reciprocal.²²

What explains the much higher levels of trust and positive affect in this one condition, if the partners' behaviors do not account for them? We believe the answer lies in the *change* in the relation that subjects in this condition, but none of the other conditions, experienced. As figure 3 shows, whereas actors in the low-power network settled into committed relations between A and B almost immediately when exchanges were negotiated, those engaged in reciprocal exchange did not. Instead, the initial level of commitment in these relations was quite low—a pattern that was potentially more profitable for the powerful As, who could receive greater benefits in the low-power networks by exchanging primarily with each other while giving to the Bs only intermittently. But this pattern was also riskier

²² This pattern is weaker and nonsignificant for feelings of commitment, primarily because the mean values in the low-power, negotiated-exchange network are higher than those for the other measures. It is likely that the high levels of behavioral commitment in this network produced stronger feelings of commitment, even though they had less effect on trust and affect.

and more uncertain. Consequently, the As gradually formed committed relations with the Bs, opting for the stable and certain benefits that the more dependent Bs provided.

This change meant that both actors in these relations experienced their early exchanges as uncertain and risky, whereas their later exchanges were stable, predictable, and mutually beneficial. The unpredictability of the early exchanges should heighten actors' awareness of risk and the potential for exploitation, thus increasing their positive evaluations of their partners' subsequent commitment. These conditions are probably closest to those envisioned by the classical theorists, in which exchange relations evolve slowly as actors first test each other's trustworthiness and gradually increase the frequency and regularity of their exchange.

CONCLUSIONS AND IMPLICATIONS

Our results offer strong support for one of the most intriguing propositions of classical exchange theory: that reciprocal forms of exchange, in which actors individually provide benefits to each other without knowing what returns they will receive, provide a more fertile ground for the development of trust than negotiated exchanges with binding agreements. As we predicted, trust in the partner and related feelings of affective commitment were significantly stronger when exchanges were reciprocal rather than negotiated. In addition, exchange behaviors that should influence whether actors come to trust or distrust their partners—the partner's behavioral commitment and the equality or inequality of their exchange—had stronger effects on trust and affect when exchanges were reciprocal.

These findings support the underlying logic of our theory: that risk is a necessary condition for the development of trust, which then depends on the partner's behavior. Although both forms of exchange entail certain kinds of ambiguity and risk, only reciprocal exchange involves the type of risk—the risk of giving without reciprocity—that requires trusting another and that allows the demonstration of trustworthiness. In negotiated exchanges, known terms and binding agreements provide "assurance" that one will not be exploited, without enabling trust.

A controlled experiment was a valuable tool for testing these predictions; it allowed us to compare reciprocal and negotiated exchanges under equivalent conditions, holding constant other factors that might affect trust, so that we could disentangle the effects of the form of exchange from other, potentially confounding factors. We physically equated the structure of the exchange networks, the relative power of the actors, the type and value of benefits exchanged, the general physical conditions under which exchanges occurred, and actors' information about all aspects of exchange except those that are part of the defining differences between

the two forms of exchange. We then statistically controlled for the frequency and equality of actors' exchanges with each other. With nothing left to vary but the form of exchange itself, we still found strong and consistent differences in trust and affective commitment.

That the form of exchange made such a difference for emergent properties like trust and affective commitment is particularly remarkable given some of the restrictions of our study. Although our experimental setting allowed a far more unambiguous test of the effects of the form of exchange than would comparisons in a natural setting, creating reciprocal and negotiated forms of exchange in the laboratory also meant that many features that typically surround reciprocal exchanges, and that serve to further differentiate them from negotiated exchanges, were absent. Unlike reciprocal exchanges in natural settings, individuals in the experimental setting never met each other or interacted face-to-face. The benefits they exchanged were as quantified as those in the negotiated exchanges, and there was no inherent "symbolism" of friendship in their nature. Delays in reciprocity were minimal; that is, exchanges that would normally take place over weeks or months in natural settings were necessarily compressed into a two-hour period, so that actors were not left in uncertainty about the other's reciprocity for very long. And the fixed benefits that we used to control the value of exchanges in the two settings meant that actors could not follow Blau's prescription of beginning with minor exchanges in which little value is involved and then gradually expanding to exchanges of greater value as partners prove themselves trustworthy. *Despite all of these limitations, our subjects still came to trust their partners and to feel positively toward them to a much greater degree in the reciprocal exchanges than in the negotiated exchanges.*

We believe these findings result from the key analytical distinctions between the two forms of exchange that the classical theorists originally identified, and that—despite limitations—we created in our laboratory settings. In reciprocal exchanges, actors choose, individually, to give to one another, without any formal assurance of reciprocity. No matter how established the relation, how predictable the other's behavior, and how long the "shadow of the future," each act of giving still remains a declaration of trust that the other will reciprocate, and each act of reciprocity confirms that trust. In contrast, the negotiation of deals, even in committed relations in which such deals are made quickly, conveys the opposite message. It is an implicit statement that knowledge of what the other will do (and, under the conditions we studied, a binding agreement to assure they will do it), is necessary before exchange can proceed. Under these conditions, trust is irrelevant and trustworthy behavior is difficult to detect.

Our work also shows the important role that behavioral commitments

play in the development of trust, particularly when exchanges are reciprocal. For both forms of exchange, the partner's commitment—the frequency of their exchange with the actor, to the exclusion of other alternatives—was more important than the equality of the exchange in subjects' evaluations of the partner's character, although it had far greater impact on evaluations of partners engaged in reciprocal exchange. Not only were relations between the inequality of exchange and evaluations of the partner relatively weak, but subjects' evaluations of their partners were not correlated with their exchange benefits in any consistent way. The subjects who earned the most money in the experiment, for example, were powerful actors in the high-power, negotiated exchange networks, yet these subjects reported the lowest levels of trust, positive affect, and feelings of commitment toward their partners of any of the powerful actors in the experiment. These findings support the emphasis in much of the literature on the importance of long-term relations and repeated interactions for developing trust (Dasgupta 1988; Lawler and Yoon 1996, 1998). Lawler and Yoon, in particular, have emphasized the importance of repeatedly and successfully completing agreements—regardless of the specific terms—for creating strong affective ties between partners in negotiated exchange relations. Our results show that such commitments have even stronger effects in reciprocal exchange relations.

Our findings also suggest that behavioral commitments may be more effective when they develop gradually. The highest levels of behavioral commitment occurred in the low-power networks, which reduced the opportunity costs of commitments for both negotiated and reciprocal exchange. But the speed with which actors formed these commitments varied markedly. In negotiated exchanges, committed relations formed early and changed little over time; in reciprocal exchanges, commitments were formed gradually, in a step-by-step process. We believe the slow transformation of the reciprocal exchange relations, from uncertainty and instability to mutual predictability and benefit, produced the particularly high levels of trust and affect observed in this condition. Further study of how the speed and process of commitments affects the development of trust is clearly warranted.

In many discussions, commitment and trust have been treated as paired processes, interrelated through either their reciprocal effects on one another (Blau 1964) or their common roots in uncertainty and risk (Kollock 1994). In contrast, our study suggests that although risk and uncertainty are implicated in the development of both behavioral commitment and trust, their relations to the two are quite different. Behavioral commitment is a means of reducing the uncertainty of both reciprocal and negotiated exchange, and it is no more likely to be used in one than in the other. A more important determinant of commitment is the structure of power

and its effects on the opportunity costs of commitment. Trust, on the other hand, does not reduce uncertainty so much as it frees actors to take the risks associated with uncertainty, risks that may lead to greater benefit. But trust can develop only in the presence of a particular kind of risk, the risk of the partner's nonreciprocity, which is more strongly associated with reciprocal than negotiated exchange.

The substantial role that both reciprocal exchanges and behavioral commitments play in the development of trust and affective ties has important implications for everyday social life. In natural settings, the clear analytical distinctions between reciprocal and negotiated exchanges that we have studied can become blurred. Many exchanges involve elements of both; even economic exchanges are heavily embedded in ongoing social relationships (DiMaggio and Louch 1998; Granovetter 1985). Numerous authors have discussed the importance of such relationships for generating trust and social capital (e.g., Coleman 1990). Many of these discussions, however, assume that the most important element in this process is the *recurrence* or *persistence* of exchange between the same actors. Our results show that not only the persistence but the *form* of social relationships matters; that is, relationships characterized by both reciprocal exchange *and* the expectation of continued interaction are particularly conducive to building trust, and they can be valuable assets in even the most institutionalized economic settings. If, instead, the shadow of the future is relatively short and exploitation is profitable, the risk inherent in reciprocal exchanges may outweigh the benefits. It is precisely for this reason that assurance structures were developed. As our results show, however, assurance has a price: a decline in trust.

The purely reciprocal exchanges that characterize many of our interactions with friends, family, and neighbors can also serve a valuable function. Through numerous experiences with specific others who behave in a trustworthy manner under conditions of risk, we may come to expect that others, with whom we have had no direct experience, will also be worthy of our trust. In this way, reciprocal exchange relations can contribute to a more generalized sense of trust in others. Establishing such a trusting environment can be a great advantage to society: individuals are free to explore new relations and take advantage of new opportunities (Yamagishi et al. 1998), and social, business, and political dealings are all facilitated.

REFERENCES

- Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.
 Blau, Peter M. 1964. *Exchange and Power in Social Life*. New York: Wiley.
 Bonacich, Phillip. 1995. "Four Kinds of Social Dilemmas within Exchange Net-

- works." *Current Research in Social Psychology* 1:1-6. <<http://www.uiowa.edu/~grpproc>>
- Budescu, David V., Ido Erev, and Rami Zwick, eds. 1999. *Games and Human Behavior*. Mahwah, N.J.: Lawrence Erlbaum.
- Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, Mass.: Harvard University Press.
- Cook, Karen S., and Richard M. Emerson. 1978. "Power, Equity and Commitment in Exchange Networks." *American Sociological Review* 43:721-39.
- Cook, Karen S., Richard M. Emerson, Mary R. Gillmore, and Toshio Yamagishi. 1983. "The Distribution of Power in Exchange Networks: Theory and Experimental Results." *American Journal of Sociology* 89:275-305.
- Dasgupta, Partha. 1988. "Trust as a Commodity." Pp. 49-72 in *Trust: Making and Breaking Cooperation Relations*, edited by Diego Gambetta. New York: Basil Blackwell.
- DiMaggio, Paul, and Hugh Louch. 1998. "Socially Embedded Consumer Transactions: For What Kinds of Purchases Do People Most Often Use Networks?" *American Sociological Review* 63:619-37.
- Ekeh, Peter P. 1974. *Social Exchange Theory: The Two Traditions*. Cambridge, Mass.: Harvard University Press.
- Emerson, Richard M. 1972a. "Exchange Theory, Part I: A Psychological Basis for Social Exchange." Pp. 38-57 in *Sociological Theories in Progress*, vol. 2. Edited by Joseph Berger, Morris Zelditch, Jr., and Bo Anderson. Boston: Houghton-Mifflin.
- . 1972b. "Exchange Theory, Part II: Exchange Relations and Networks." Pp. 58-87 in *Sociological Theories in Progress*, vol. 2. Edited by Joseph Berger, Morris Zelditch, Jr., and Bo Anderson. Boston: Houghton-Mifflin.
- . 1981. "Social Exchange Theory." Pp. 30-65 in *Social Psychology: Sociological Perspectives*, edited by Morris Rosenberg and Ralph H. Turner. New York: Basic Books.
- Friedkin, Noah. 1993. "An Expected Value Model of Social Exchange Outcomes." Pp. 163-93 in *Advances in Group Processes*, vol. 10. Edited by Edward J. Lawler, Barry Markovsky, Karen Heimer, and Jodi O'Brien. Greenwich, Conn.: JAI Press.
- Gambetta, Diego. 1988. "Can We Trust Trust?" Pp. 213-37 in *Trust: Making and Breaking Cooperation Relations*, edited by Diego Gambetta. New York: Basil Blackwell.
- Granovetter, Mark. 1985. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91:481-510.
- Hardin, Russell. 1991. "Trusting Persons, Trusting Institutions." Pp. 185-209 in *Strategy and Choice*, edited by Richard J. Zeckhauser. Cambridge, Mass.: MIT Press.
- Heckathorn, Douglas D. 1985. "Power and Trust in Social Exchange." Pp. 143-67 in *Advances in Group Processes*, vol. 2. Edited by Edward J. Lawler. Greenwich, Conn.: JAI.
- Kagel, John H., and Alvin E. Roth, eds. 1995. *The Handbook of Experimental Economics*. Princeton, N.J.: Princeton University Press.
- Kelley, Harold H., and John W. Thibaut. 1978. *Interpersonal Relations: A Theory of Interdependence*. New York: Wiley.
- Kollock, Peter. 1994. "The Emergence of Exchange Structures: An Experimental Study of Uncertainty, Commitment, and Trust." *American Journal of Sociology* 100:313-45.
- Lawler, Edward J. 1992. "Power Processes in Bargaining." *Sociological Quarterly* 33: 17-34.
- Lawler, Edward J., and Jeongkoo Yoon. 1996. "Commitment in Exchange Relations: Test of a Theory of Relational Cohesion." *American Sociological Review* 61:89-108.

- . 1998. "Network Structure and Emotion in Exchange Relations." *American Sociological Review* 63:871–94.
- Lévi-Strauss, Claude. 1969. *The Elementary Structures of Kinship*, rev. ed. Boston: Beacon.
- Lewis, J. David, and Andrew Weigert. 1985. "Trust as a Social Reality." *Social Forces* 63:967–85.
- Luhmann, Niklas. 1979. *Trust and Power*. Chichester: Wiley.
- Macy, Michael W., and John Skvoretz. 1998. "The Evolution of Trust and Cooperation between Strangers." *American Sociological Review* 63:638–60.
- Markovsky, Barry, John Skvoretz, David Willer, Michael J. Lovaglia, and Jeffrey Erger. 1993. "The Seeds of Weak Power: Extending Network Exchange Theory." *American Sociological Review* 58:197–209.
- Molm, Linda D. 1990. "Structure, Action and Outcomes: The Dynamics of Power in Exchange Relations." *American Sociological Review* 55:427–47.
- . 1997. *Coercive Power in Social Exchange*. Cambridge: Cambridge University Press.
- Molm, Linda D., and Karen S. Cook. 1995. "Social Exchange and Exchange Networks." Pp. 209–35 in *Sociological Perspectives on Social Psychology*, edited by Karen S. Cook, Gary Alan Fine, and James S. House. Boston: Allyn & Bacon.
- Molm, Linda D., Gretchen Peterson, and Nobuyuki Takahashi. 1999. "Power in Negotiated and Reciprocal Exchange." *American Sociological Review*, vol. 64, no. 6, in press.
- Putnam, Robert D. 1995. "Bowling Alone: America's Declining Social Capital." *Journal of Democracy* 6:65–78.
- Schelling, Thomas C. 1960. *The Strategy of Conflict*. Cambridge, Mass.: Harvard University Press.
- Skvoretz, John, and David Willer. 1993. "Exclusion and Power: A Test of Four Theories of Power in Exchange Networks." *American Sociological Review* 58:801–18.
- Tallman, Irving, Louis Gray, and Robert Leik. 1991. "Decisions, Dependency, and Commitment: An Exchange-Based Theory of Group Development." Pp. 227–57 in *Advances in Group Processes*, vol. 9. Edited by Edward J. Lawler, Barry Markovsky, Cecilia Ridgeway, and Henry Walker. Greenwich, Conn.: JAI Press.
- Yamagishi, Toshio, Karen S. Cook, and Motoki Watabe. 1998. "Uncertainty, Trust and Commitment Formation in the United States and Japan." *American Journal of Sociology* 104:165–94.
- Yamagishi, Toshio, and M. Yamagishi. 1994. "Trust and Commitment in the United States and Japan." *Motivation and Emotion* 18:129–66.

The Mosaic Moment: An Early Modernist Critique of Modernist Theories of Nationalism¹

Philip S. Gorski
University of Wisconsin, Madison

Most theorists of nationalism are modernists: they regard nationalism as an essentially modern phenomenon. This article takes issue with the modernist position. Drawing on primary and secondary evidence from the Netherlands, England, and other early modern polities, it documents the existence of movements and ideologies that must be classified as national and nationalist by the modernists' own criteria. It is then argued that some nationalist discourses had medieval roots and that they were no less nationalistic than the nationalisms of the French Revolution. In the conclusion, the theoretical premises of the modernist position are subjected to critical examination.

Die Form ist flüssig; der Sinn ist es aber noch mehr.
Friedrich Nietzsche, *Zur Genealogie der Moral*

Most theorists of nationalism are modernists:² that is, they believe that nationalism is an essentially modern phenomenon. Modernists do not claim that national identity is a modern phenomenon; they usually concede the existence of some form of premodern "national consciousness" or

¹ Earlier versions of this article were presented at the annual meeting of the Social Science History Association (New Orleans, 1997), the Wilder House Workshop on Nations and Nationalism (University of Chicago, 1998), the Institute for Research in the Humanities (University of Wisconsin), and the Cultural Pluralism Circle (University of Wisconsin). I would like to thank Meyer Kestnbaum, David Laitin, Lisa Wedeen, Paul Boyer, Crawford Young, and other seminar participants for their stimulating and useful feedback. I would also like to thank Stephen Bunker, Susanne Desan, Robert Kingdon, Thomas Safley, and the reviewers at *AJS* for helpful comments on various drafts of this article. Direct correspondence to Philip Gorski, Department of Sociology, University of Wisconsin, 1180 Observatory Drive, Madison, Wisconsin 53706. E-mail: pgorski@ssc.wisc.edu

² I borrow this term from Anthony Smith (1994) but use it somewhat differently than he does. For Smith, modernists are those who deny the existence of premodern national identity. For me, modernists are those who deny the existence of premodern nationalism.

“national identity.” But they invariably insist that national-ism, *strictu sensu*, appeared no earlier than the late 18th century. This is because they generally view nationalism as the product of the democratic and industrial revolutions that began at this time.

The purpose of this article is to call the modernist position into question. In particular, I will attempt to show that premodern “national consciousness” was more like “modern nationalism” than the modernists have allowed, and, indeed, that no sharp line can be drawn between the two. To put it more pointedly, my argument is that clear examples of nationalism can be found in the early modern period, and perhaps even earlier; that the primary catalyst for this nationalist upsurge was the Protestant Reformation and the political revolutions that followed in its wake; and that this upsurge was indeed an *upsurge*, rather than the “birth,” of nationalism.

The main exhibit in my case against the modernists is the early modern Netherlands. Drawing on the works of a number of Dutch historians (Geyl [1930] 1975, 1934; Smitskamp 1947; Huisman 1983; Groenhuis 1977; Schöffner 1981; Schama 1988), and on a variety of primary sources, including political pamphlets, popular songs, commemorative coins, and ritual prayers, I will show that political conflict and debate in the early modern Netherlands led to the generation of two main myths or narratives about the origins of the Dutch Republic: a biblical narrative, grounded in the Old Testament, which portrayed the Netherlanders as a chosen people or elect nation bound to God by a holy and sacred covenant, and the Netherlands as a new Israel or new Jerusalem; and a classical narrative that portrayed the Netherlanders as the descendants of the ancient Batavians, a free and peace-loving people that had always resisted imperial tyranny. And I will argue that these mythologies or narratives, and the ideologies and movements they inspired, must be characterized as nationalist—even by the definitions and standards employed by the modernists.

I will buttress these claims with comparative and historical evidence that shows (1) that the Hebraic and classical mythologies were not distinctive to the Netherlands, but can also be observed in a number of other early modern polities, most notably England, (2) that examples of the Hebraic and classical mythologies can also be found in the Middle Ages as well, and (3) that the content and scope of early modern nationalism did not differ dramatically from that of French revolutionary nationalism, the paradigmatic case of modern nationalism. In the conclusion, finally, I will present the outlines of a postmodernist theory of nationalism. Before plunging into the historical evidence in greater depth, it will perhaps be useful to review the various versions of the modernist position, and my critique of them, in somewhat greater detail.

THE MODERNIST POSITION: A CRITICAL OVERVIEW

Theorists of nationalism disagree strongly about what nationalism is and why it came into being. Some see nationalism as an ideology (Hroch 1985; Brass 1991), others as an identity (Anderson 1992; Balibar 1991), and still others as a movement (Breuilly 1982). The explanations are as diverse as the definitions. Some attribute the rise of nationalism to the advent of industrialization (e.g., Gellner 1983), others trace it to the French Revolution or German romanticism (Hobsbawm 1990; Kemiläinen 1964; Kedourie 1993), and still others to new forms of communication (Deutsch 1966).

Given these disagreements about the what and the why of nationalism, one might expect similar disagreements about the when question. But this is not the case. In fact, with only a few exceptions (Smith 1991, 1994; Greenfeld 1992), most scholars of the subject seem to agree that nationalism is a distinctly modern phenomenon that first appeared sometime during the late 18th century.

This consensus is all the more surprising given that there are at least two bodies of evidence that speak against it. The first is philological. There is a fairly extensive literature that focuses on the word "nation" and kindred terms such as "people" and "state" in both classical Greek and Latin as well as the various European vernaculars (see, e.g., Volk [1972–97] and the references therein and also Heissenbüttel [1920], Kahl [1978], and Eichenberger [1991]).³ What this literature shows is that these categories have long been a part of Western political discourse and that they have always been closely related to one another in meaning. In other words, it suggests that nationalism qua ideology antedates the modern era.

The second body of evidence is historical. There is now a rather substantial literature on the subject of medieval and early modern nationalism (see esp. the relevant chapters in the following edited collections: Bradshaw and Roberts 1998; Ranum 1975; Forde, Johnson, and Murray 1995; Ecclesiastical History Society 1982; Berding 1994; Giesen 1991; Beumann and Schröder 1978; Beumann 1983; Dann 1986; Ehlers 1989; and see also Beaune 1985; Schmugge 1982; O'Brien 1988; Hastings 1997; Reynolds 1997). What this literature clearly shows is that a wide range of medieval and early modern authors talked about nations and peoples in a wide variety of scholarly and political contexts and in ways that are quite similar, if not wholly identical, to modern usage, and that these discourses played an important role in political legitimation and mobiliza-

³ Those scholars of nationalism who have taken the philological evidence most seriously (i.e., Smith and Greenfeld) have tended to rely on two overly simplistic and schematic works (Zernatto 1944; Kemiläinen 1964).

tion. In other words, it implies that nationalism qua identity and movement may also be older than most theorists of nationalism believe.

To the degree that theorists of nationalism have engaged with this evidence—and most, frankly, have not—they have responded in one of two ways. Some—the majority—concede the existence of some form of premodern national consciousness but argue that it did not constitute true or full-blown nationalism. Others—a minority—concede the existence of some form of premodern nationalism but argue that it was basically identical to modern nationalism, or should be treated as such. Thus, both deny the existence of an early modern nationalism, albeit for rather different reasons. Let us examine these views in greater detail, beginning with the majority view.

The majority view has two main variants: one focuses on the *content* of nationalism; the other on the *scope* of nationalism. There are two subvariants of the content argument. The first focuses on nationalism qua ideology. John Breuilly's (1982; see also Gellner 1983; Smith 1994, pp. 379–80; 1995, pp. 25–26) work provides a good example of this approach. Breuilly (1982, p. 3) argues that “a nationalist argument is a political doctrine built upon three basic assertions: (1) There exists a nation with an explicit and peculiar character. (2) The interests and values of this nation take priority over all other interests and values. (3) The nation must be as independent as possible.” In his view, the “nationalist argument” did not appear in its fully developed form until the era of the French Revolution. The second subvariant of the meaning argument focuses on nationalism qua category. This is the approach taken by Eric Hobsbawm (1990; see also Calhoun 1997, p. 9). Hobsbawm argues that nationalism is founded on the equation “nation = people = state” (1990, p. 12). In his view, this “sense of the word is no older than the eighteenth century, give or take the odd predecessor” (1990, p. 3).

The second version of the majority view, to recall, is focused on the *scope* of national discourse. It also has two subvariants. The first focuses on the *social* scope of nationalism. Here, the usual procedure is to argue that premodern national consciousness was limited to certain elite groups (e.g., intellectuals), subordinated to other identities (e.g., to one's religion, class, or region), or confined to relatively small political units (e.g., cities and regions), and thus not fully national in one sense or another. This argument has fallen out of favor among sociologists, but it was a stock-in-trade of the first-generation theorists of nationalism (e.g., Kohn [1944] 1967, p. 12; Hayes 1931, p. 4; see also Armstrong 1982) and can still be found in the works of some historians (e.g., Weber 1976; Sahlins 1989). The second and more common subvariant of the scope argument concerns *political* scope. Here, the usual procedure is to argue that premodern na-

tional consciousness did not give rise to political movements or that such movements were not truly or fully nationalist in some sense. Once again, John Breuilly's work (1982) provides a particularly good example. Breuilly openly acknowledges the existence of "national opposition movements" during the early modern period (i.e., revolutionary movements that deployed nationalistic arguments). But he insists that these movements were not fully nationalist because "the idea of the nation, if it appeared at all, was subordinated to religious and monarchical principles" (1982, p. 45). The first truly nationalist movements, he reiterates, emerged during the French Revolution.⁴

Now for the minority view—the view that there *was* such a thing as premodern nationalism but that it was essentially identical to modern nationalism, or should be treated as such. To my knowledge, this view has been advanced by only two theorists: Liah Greenfeld and Anthony D. Smith. Greenfeld's argument is that modern nationalism was really an early modern phenomenon. More specifically, she contends that the origins of modern nationalism are to be found, not in revolutionary France, but in Reformation England (1992, p. 6; for a similar argument, see also Hastings [1997, p. 4]). For it was there, she says, that the word nation came to mean "a sovereign people," the central meaning that (she contends) it has retained to this day.

Smith's argument is more equivocal. This is not surprising: Smith spent over two decades defending the majority view (see esp. Smith 1986) and has become critical of it only in recent years (see Smith 1991, 1994, 1995). His critique of the modernist position rests on the distinction between nationalism qua ideology and nationalism qua identity. If one sees nationalism as an ideology, he says, then the modernists are right, because in his view, the ideology of nationalism really is no older than the French Revolution. If one sees nationalism as an identity, however, then the modernists are wrong, since one can find numerous examples of national identity that antedate the French Revolution. Indeed, Smith goes so far as to suggest that the ancient Israelites possessed national identity (see also Grosby 1991).

So who is right? The majority? Or the minority? Before attempting a preliminary answer to this question, let us quickly review the empirical claims on which each position rests. Advocates of the majority position propose various arguments: (1) The modern meaning of the word "nation," summarized in the equation "nation = people = state," does not antedate the French Revolution. (2) The modern ideology of nationalism, which

⁴ Needless to say, the meaning and scope arguments are not necessarily exclusive of one another. As the example of Breuilly's work clearly illustrates, many theorists of the majority persuasion draw on some variant of each argument.

claims that the world is divided into nations, that these nations are sovereign, and that sovereignty requires statehood, also does not antedate the French Revolution. (3) national consciousness or identity did not become generalized until the French Revolution. (4) there were no truly or fully nationalist movements until the French Revolution. The minority argument consists of two distinct but opposed counterclaims: (1) Modern nationalism is an early modern phenomenon that begins in England. (2) Nationalist ideology is a modern phenomenon that begins with the French Revolution, but national identity is a perennial phenomenon that can be traced back to antiquity.

Generally speaking, I am inclined toward the minority view, and in what follows, I will try to show that all four of the major claims on which the majority position rests are exaggerated. In particular, I will contend that (1) the early modern and modern meanings of the word *nation* are not so different as Hobsbawm and others have suggested, (2) early modern and modern ideologies of nationalism were not so different as Breuilly and others have suggested, (3) early modern and modern national consciousness were not so different in scope as Kohn and others have suggested, and (4) the national opposition movements of the Reformation Age were surprisingly nationalist, pace Smith. In other words, I will try to show that *some instances of early modern national consciousness must be counted as instances of full-blown, modern nationalism by the very criteria set forth by the modernists.*

But this does not mean that I completely endorse the minority view. On the contrary, I have serious reservations about it. In fact, I believe that Greenfeld's and Smith's arguments are seriously deficient in several respects. As I will try to show, Greenfeld's claim that modern nationalism originated in early modern England is both empirically and interpretively problematic: empirically, insofar as the kind of nationalism that Greenfeld discovers in early modern England—what I will call holy or Hebraic nationalism—can also be observed in the early modern Netherlands and in a number of other early modern polities as well; and interpretively, insofar as it elides and downplays some very important differences between modern and early modern nationalisms. Rather than arguing that modern nationalism was an early modern phenomenon, I would like to argue that Hebraic nationalism was a phenomenon unto itself, which was different from what came before and after it. Equally problematic, in my view, is Smith's claim that "national identity" is perennial, and "nationalist ideology" modern, not only because nationalist ideology is *not* modern, but because national identity and national ideology tend to arise in tandem, rather than in sequence. That, at least, is the conclusion that is suggested by my reading of the historical record.

In order to substantiate these claims, it is now necessary to examine

the discourses and narratives of early modern nationalism, and the ideologies and movements to which they gave rise, in considerably greater detail. I begin with the discussion of the Netherlands, which makes up the empirical core of this article.

MODERN NATIONALISM IN THE EARLY MODERN NETHERLANDS

To understand the rise of Dutch nationalism, one must understand something of Dutch history and of European history more generally. The 16th and 17th centuries were a time of great turmoil in Western Europe, marked by an incendiary mixture of constitutional conflict (princes vs. parliaments, patricians vs. burghers, lords vs. peasants) and confessional strife (Catholics vs. Protestants, Lutherans vs. Calvinists, churches vs. sects), which catalyzed a century and a half of warfare and rebellion throughout the British Isles and across most of the continent. The Low Countries, which encompassed most of present-day Holland and Belgium, were not immune (Groenveld et al. 1991; Israel 1995). On the contrary, they were severely affected. They experienced constitutional conflict generated by the protoabsolutist policies of the Habsburgs and Dutch resistance to them. And they experienced confessional conflict generated by the anti-Protestant policies of the Habsburgs and the rise of Dutch Calvinism. The combination of these two conflicts sparked the Dutch Revolt against Spain (ca. 1555–1609) and, eventually, the division of the Low Countries—into a Calvinist republic (the Dutch Republic), in the North, and a Catholic principality (the Spanish Netherlands), in the South (Parker 1985).

Nor did this bring an end to the turmoil. To be sure, the Twelve Years Truce (1609–21) between Habsburg Spain and the Northern Netherlands brought a (temporary) halt to the fighting. But it also unleashed a bitter feud within the Dutch Republic, commonly known as the “Religious Quarrels” (*Twisten van Religie*; Nobbs 1938; Tex 1973; Deursen 1991). On their face, the Religious Quarrels were just that—religious quarrels regarding Calvin’s doctrine of predestination and the government of the Reformed Church. They pitted defenders of Calvinist orthodoxy (Precisionists) and church autonomy (Consistorians) against advocates of free will theology (Arminians) and secular control of the Church (Erastians). But the Religious Quarrels had political undertones, and they soon became entangled with constitutional quarrels regarding the true locus of sovereignty and the proper distribution of authority within the Republic. On the one side stood the province of Holland and its patrician rulers, who argued that sovereignty resided in the individual provinces and that legislative authority belonged to the provincial parliaments or states. On the other side stood the peripheral provinces and the Dutch stadtholders, who argued that sovereignty resided in the national parliament or States

General and that the states should share power with the stadtholders. Since the ruling class of Holland generally favored the Arminians, and the House of Orange usually sided with the Calvinists, the religious and political conflicts gradually became aligned with one another. The result was a religiopolitical showdown that pushed the young republic to the brink of civil war in 1617. This confrontation ended in a victory for the Calvinists and Orangists. But it was short-lived. By the late 1620s, the Arminians and the States Party (i.e., the patricians of Holland and their allies) were once again on the ascendant. Over the next century and a half, battles between these two parties flared up at fairly regular intervals—in 1650, when the young stadtholder, William II, staged a coup against the Arminians, only to succumb to smallpox several months later; in 1672, when the French and the English invaded the Netherlands and the province of Holland was compelled to place William III in charge of the armed forces; and several more times in the 18th century (Rowen 1988; Roorda 1961; Bijl 1981; Groenveld 1967, 1990). It is against the background of these events—the Dutch Revolt, the Religious Quarrels, and the partisan conflicts that followed—that the emergence and development of Dutch nationalism must be seen.

Early modern Dutch nationalism had several different strands. The first and most important was spun out of the symbols and stories of the Old Testament. It portrayed the Dutch as a chosen people who had formed a sacred covenant with God, a people who had been blessed by God with sanctity, liberty, and riches, and who would continue to be thus blessed so long as they upheld God's laws and commandments. Following Connor O'Brien (1988; see also Walzer 1985), I will refer to this strand as holy or Hebraic nationalism. The second strand drew on the symbols and stories of classical antiquity and especially on Tacitus's *Germania*. It portrayed the people of Holland, and sometimes of the Netherlands more generally,⁵ as the descendants and heirs of the ancient Batavians, a freedom-loving people who had valiantly fought off the domination of the Romans and the Spaniards (Schöffner 1981; Schama 1988, pp. 76–81). The third strand was probably inspired by the mythology of the French monarchy. It portrayed the princes of Orange as the divinely appointed protectors of the Netherlands. The fourth strand, finally, anticipated the ideology of the French Revolution. It depicted the Dutch *menu peuple* as the true founts of sovereignty and the pillars of the nation.

⁵ In English, the Netherlands as a whole are sometimes referred to as Holland, and the Northern Netherlands are sometimes referred to as the Low Countries. I will follow the Dutch usage, using the term "Holland" only to refer to the province of Holland, and the term Low Countries only to refer to the Northern and Southern Netherlands—the present-day Netherlands and Belgium—as a whole.

These strands emerged at different moments, evolved over time, and were intertwined in various and changing ways. Hebraic nationalism emerged during the Dutch Revolt, classical nationalism during the Quarrels, and monarchical and republican nationalisms during the partisan conflicts of the 1660s and 1670s. Hebraic and classical nationalism were initially opposed, while monarchical and republican nationalism went hand-in-hand; later, these relationships were reversed.

The key point for the present analysis, however, is not the complex ways in which these various discourses were related to one another, but the degree to which they resembled modern nationalism as defined by the modernists. As we will see directly, early modern Dutch nationalism was national in scope almost from the outset. And by the mid-17th century, it was quite nationalist in content as well.

The Dutch Revolt and the Invention of Hebraic Nationalism

It is difficult to know precisely when the leaders of the Dutch Revolt first came to see their struggle in Hebraic terms, but there can be little doubt that the shift had occurred by September 1577, when William of Orange made a triumphal entry into Brussels, because the ceremonies that accompanied his arrival were rife with Hebraic analogies (Knuttel 320).⁶ Before passing through the city gates, he and his entourage were rowed down the River Scheldt past a series of floats. The first showed David holding Goliath's head in one hand and a sword called "the Strength of God" (*de cracht Godts*) in the other (see fig. 1). An allegorical figure dressed in the garb of a Dutch burgher explained its significance as follows:

Mark well: David being a shepherd of sheep
Was chosen by God for his humility
To slay Goliath great
Who had sworn the death of Israel. (Knuttel 320, p. 27)

The message was clear: just as the young David had slain the towering Goliath, so William had defeated the overmighty king of Spain. The second float showed Moses holding a rod in his hand called "the Power of God" (*de macht Godts*; see fig. 2). Opposite him were the Children of Israel, looking through the bars in the door of a dungeon called "Slavery in Egypt." Moses then struck the bars with his rod and liberated them (Knuttel 320, p. 30). Here, an analogy was being drawn between two great

⁶ In this and subsequent citations, "Knuttel" refers to the Knuttel Pamphlet Collection, housed at the Royal Library (*Koninklijke Bibliotheek*) in The Hague. The pamphlets are cataloged in Knuttel (1889–1920), and the numbers in the citations refer to the numbers therein. These pamphlets are also available on microfiche as Dutch Pamphlets (1979–82).

FIG. 1.—David and *de cracht* Gods

liberators—Moses and William. The third float portrayed Joseph holding a staff, which was painted white, orange, and blue, the colors of the House of Orange-Nassau, and called “the Wisdom of God.” Joseph was bound by a chain labeled “oppression” and held by a figure named “discord.” Using the staff, he knocked down discord and freed Jacob and all his lineage from the hands of slavery. The implication was clear: by forging national unity, William had won freedom from Spain (Knuttel 320, p. 32).

Taken together, the three *tableaux* form a complex allegory of the Dutch Revolt up to 1577: the beheading of the giant warrior by the poorly armed upstart (the invasion of 1572); the deliverance of the Jewish slaves from the cruel Pharaoh (the establishment of an independent republic in the North in 1573); and the restoration of internal concord among the tribes of Israel (the Pacification of Gent in 1576)—an effect, it should be noted, that was achieved only by reversing the narrative order of the biblical events.



FIG. 2.—Moses and *de macht Gods*

Though the usage of biblical symbolism in the joyous entry of 1577 was particularly rich, it was by no means unique. Comparisons between the events of the Dutch Revolt and the stories of the Old Testament were also a common device in the patriotic fight songs of the late 16th and early 17th centuries, the so-called "beggars songs" (*Geuzenliederen*). Many of these songs compared William of Orange to Moses and David, and the king of Spain to the Pharaoh and Saul (Kuiper 1924–25, nos. 45, 65, 129; see also Duyse 1965, vol. 2, no. 432). The Dutch people, meanwhile, were variously described as "God's elect" (*Gods uutvercoren*), "God's people" (*Godts volck*), or "little lambs" (*schaepkens*), while the Spaniards were excoriated as a "foreign nation" (*vreemde nacy*), distinguished by its "arrogance" (*hoochmoet*) and "cruelty" (*wreedheid*; Kuiper 1924–25, nos. 38, 45).

Many of those songs that invoked Hebraic imagery seem to have derived from the Calvinist milieu. But if the Calvinists were the first to

invoke Hebraic imagery, they were by no means the last. Such images were a recurring motif in the memorial pennies minted by many civic and provincial governing bodies during the revolt. One of the earliest examples is a coin issued to commemorate the end of the Spanish siege on Leiden (see fig. 3). The front shows an avenging angel driving the king of Assyria, Sennach'erib, from Jerusalem (2 Kings 19), while the back depicts the Spanish siege of Leiden. The inscription reads: "Like Sennach'erib in Jerusalem, the Spaniards fled Leiden by night" (Loon 1723–31, vol. 1, pp. 193–95). Another penny minted that year, probably by the States of Holland, shows a cornerstone on one side and the Hebrew word for Lord (Jehovah) with a radiant cloud in the background on the other (Loon 1723–31, vol. 1, pp. 202–3). A third penny, issued in 1576, contains a further elaboration of this theme. The front shows a mounted soldier in full battle dress, and the back depicts the cloud motif—a hand descending from a radiant cloud inscribed with the Hebrew "Jehovah"—above the head of a praying man with a Dutch cityscape in the background (Loon 1723–31, vol. 1, p. 228). The cloud and hand motif was used as a symbol of divine providence for over half a century and was employed with great frequency on coins commemorating a military victory. In one coin minted shortly after the sinking of the Spanish Armada, the cloud motif had a meaning at once figurative and literal. The frontispiece shows the Spanish fleet in full disarray, with the cloud motif above (see fig. 4). The inscription reads: "The Lord blew and they were scattered" (Loon 1723–31, vol. 1, p. 392).

Since minting rights were generally limited to political bodies of one kind or another in the Dutch Republic, at least two inferences can be drawn from the usage of Hebraic imagery in Dutch coins. First, a significant faction of the political elite must have espoused the Calvinist view of the Dutch Revolt as a struggle for religious freedom, be it sincerely or cynically. Second, at least some of those who portrayed the revolt as a struggle for political liberty also portrayed it as a struggle for religious freedom, whether out of conviction or calculation. So while the biblical interpretation of the Dutch Revolt was most likely introduced by the Calvinists, it was soon adopted by members of the regent class.

No doubt, the circulation of these coins was limited, their influence small. But this cannot be said of the *biddagsbrieven*, the official proclamations that were read on days of public prayer and issued by the civil authorities. The purpose of the fast and prayer days was to ward off worldly dangers and secure divine protection, and in times of great peril or distress, local or national leaders proclaimed them in great number.⁷ The

⁷ For example, the magistrate of Leiden called no less than nine during the year of the Spanish siege (listed in Kist 1848–49, vol. 2, nos. 4–12).



FIG. 3.—Commemorative coins



FIG. 4.—Coin commemorating the sinking of the Spanish Armada

important point, for the present analysis, is that the idiom of these proclamations was distinctly prophetic in tone. The God they addressed was the God of the Old Testament, “a jealous and angry God” who defends his people from their enemies but abandons them when they “murmur” against him. In a *biddagsbrief* issued in November of 1575, for instance, William of Orange proclaims that “His Excellency, following in the footsteps of Christian Princes, who, in times of danger and distress have sought refuge in the Almighty God, and, together with their people, have humbled themselves before His almighty hand and have repented and turned away from their previous and sinful lives . . . [knowing that] God

has never left *his people* in their moment of need, but has always stood by them and delivered them" (Kist 1848–49, vol. 2, no. 18; emphasis added).

Another *biddagsbrief*, issued by the States General in February of 1580, alluded to the Exodus saga even more explicitly, importuning that "it please God to turn away the horrible plagues, the great destruction and the long-lasting war from these lands and break the power and the force of their enemies and *liberate* these lands and their good inhabitants from all which leads to their ruin and from this *accursed and eternal slavery*" (Kist 1848–49, vol. 2, no. 35; emphasis added). Hebraic imagery, then, was not only a part of official symbolism in the Dutch Republic, but a central element of public ritual as well.

Examples from the Old Testament were also common in tracts written by political propagandists and republican theorists, who used them to elaborate and defend a constitutionalist theory of government. For example, in an anonymous pamphlet written in 1576 and addressed to the States General, the author pointed out that "God became very angry, when His people of Israel, unsatisfied with a legal government, wanted to have Kings, just like the pagans" and argued that "the community is not created and ordained for the Kings' sake, but the Kings for the sake of the community" (Gelderen 1993, p. 104). The ideal government, the author implies, is an aristocratic "government of the best," like the one that ruled over Israel in the time of the judges. Writing in 1582, a second pamphleteer cited the example of the Old Testament covenant to argue that "Kings are bound by oath in two respects, first to God and then to the people"; that the king is obliged "to govern in such a way that he permits the people to serve God purely and to observe his law"; and that the people owe him their obedience only so long as he does so (Gelderen 1993, p. 193). Here, the idea of the double covenant provides the basis for a theory of political resistance—and the point of departure for an incipient notion of social contract. Such examples could easily be multiplied.

While orthodox Calvinists were certainly the most fervent disciples of the Hebraic ideology, they were by no means the only ones. Echoes of the New Israel myth may be found in the writings of moderate Protestants (Arminians), such as Johannes Uytenbogaert (Knuttel 2044, pp. 16, 19), Erastian humanists (libertines), like P. C. Hooft (1899–1900, vol. 1, p. 104; Hooft 1677, p. 3), and even devout Catholics, such as Joost van Vondel, who felt that the parallels between the slavery of the Israelites and the oppression of the Netherlanders substantiated "that in truth, one can hardly tell the one from the other" (Vondel 1937, p. 6).

Based on the foregoing evidence, it seems safe to conclude that Hebraic nationalism was quite national in scope, insofar as its use cross-cut the major social cleavages of class, confession, and region, and insofar as it was deployed in the service of a national liberation movement. Still, it

did not yet use the term “nation” and its kindred categories in a fully modern way, nor did it have that aggressively programmatic quality that many people associate with nationalism. As we will see, that would change. But before examining the development of Hebraic nationalism in further detail, let us first examine its younger sibling, and sometime rival: Batavian nationalism.

The Religious Quarrels and the Invention of Batavian Nationalism

Today, the Dutch humanist Hugo De Groot (Grotius) is best known for his writings on natural law. During the 17th century, however, the average Netherlander would probably have been more familiar with his *Treatise on the Antiquity of the Batavian Republic*. Grotius was not the first early modern author to write about the ancient Batavians, the historical forebears of the modern-day Hollanders (Kampinga [1917] 1980; Tilmans 1992). But it was the publication of Grotius's *Treatise* in 1610 that first attracted public attention to the ancient Batavians. Drawing primarily on Tacitus, Grotius argued that the Batavians had never been subjugated to monarchical or princely authority, and that Batavia had always been ruled by an aristocratic “government of the best.” Indeed, he contended that there was an essential continuity between the Dutch and Batavian Republics (Knuttel 1735, p. 5). In point of fact, the continuities were actually quite weak, as Grotius himself later recognized. But if Grotius's reading of *Germania* was somewhat biased, his sense of timing was impeccable.⁸ Members of the patriciate were put off by the theocratic pretensions of the Calvinist preachers and uncomfortable with the revolutionary roots of the Dutch Republic. Thus, they were highly receptive to Grotius's claim that the republic was an old state, not a new one, and that its origins were essentially secular. And if sales figures are any indication, the appeal of Grotius's tract must have extended to the broader citizenry as well, because it was reprinted no less than 11 times during the 17th century and at least four times during the 18th century (Molewijk 1988, p. 19). By the standards of the day, then, *Antiquity* was a huge best-seller.

While Grotius's *Antiquity* was probably the most influential version of the Batavian narrative, it was by no means the only one. There were also several stage versions (Muller 1931), such as P. C. Hooft's *Baeto* ([1626] 1954), which described the Batavians as “an able nation, which shall last through all eternity” (Hooft 1954, ll. 1429–30). While Hooft's play was not particularly successful, its eponymous hero, Baeto, did find his way

⁸ In fact, there is some evidence that suggests Grotius actually wrote the *Treatise* in 1601 and intentionally delayed its publication until after the conclusion of the Twelve Years' Truce.

into a number of popular histories that enjoyed wide circulations (Schöffner 1981, p. 98 and n. 30). The Batavian myth also found its way into official representations of the republic. In 1612, for example, just two years after the publication of Grotius's tract, the burgomasters of The Hague bought a series of 36 etchings depicting the Batavian rebellion against Rome. And the following year, the States of Holland commissioned a series of 12 paintings on the same subject, which they then hung in their meeting rooms (Waal 1952, vol. 1, pp. 210–15).

In terms of its content, the Batavian mythology was every bit as nationalistic as the Hebraic mythology—perhaps even more so. Like the Exodus saga, the Batavian narrative assumed, at least implicitly, that political sovereignty was a precondition of true nationhood. Similarly, it assumed that the roots of the nation were not only contractual, but also ethnic and cultural. In contrast to the Hebraicists, however, who tended to favor the term “people,” the Batavians were much more apt to employ the word “nation.” Indeed, it seems probable that they were largely responsible for bringing this Latinism into wider circulation.

In terms of its scope, however, the Batavian myth was undoubtedly less national than its Hebraic pendant, at least at this point in time. For while the New Israel of the Calvinists included all of the Northern Netherlands, and sometimes even the Southern Netherlands, the borders of Grotius's and Hooft's Batavia did not extend beyond the province of Holland. Moreover, the evidence suggests that the social scope of the Batavian myth was also somewhat narrower. Apart from learned treatises and public buildings, one finds few examples of it—a stark contrast to the pandemic character of Hebraic imagery during the early 17th century. This has led some Dutch historians to conclude that the Hebraic mythology was a regional or even antinational ideology. But as we will see below, this overstates the case. There were others who espoused a broader understanding of Batavia, which encompassed more of the Netherlands, and it was this understanding, rather than Grotius's and Hooft's, which became the most widely held one.

Partisan Conflict and Nationalist Ideology: From the “Further Reformation” to the “Year of Disaster”

As of 1620, Dutch national consciousness had not yet evolved into a full-blown nationalism of the modern(ist) variety. While the Hebraic myth was quite national in scope, it was not entirely nationalistic in content. For the Batavian myth, the reverse obtained: it was more completely nationalistic in its content, but it was not entirely national in its scope. By 1670, this had changed. The two myths had become fully national(istic) in scope and content and had become increasingly fused with (rather than

opposed to) one another. What is more, two new myths of the nation had come into being, one focusing on the princes of Orange, the other on the Dutch people. The stimulus for these developments were the rise of Dutch Puritanism and the fire of partisan conflict.

During the Dutch Revolt, the analogy between ancient Israel and the Netherlands had been used mainly in a retrospective and descriptive way, as a lens through which to view the past and the present. During and after the Religious Quarrels, by contrast, the Hebraic analogy was also used in a prospective and prescriptive way, as a sort of program for shaping the future.

The Hebraicist program, if we may call it that,⁹ consisted of three, inter-related goals, whose overall aim was to maintain and strengthen the Dutch nation: (1) the imposition of religious conformity and unity (i.e., the suppression of all nonorthodox and non-Calvinist groups); (2) the enforcement of a strict social and moral discipline (e.g., the implementation of sumptuary laws and compulsory observation of the Sabbath); and (3) an autonomous Church and a strong, central government.

The basic assumptions underlying this program are derived from the Pentateuch. During the revolt, it was argued, the people of the Netherlands had made a sacred covenant with God in which they agreed to uphold his laws and observe his will; in return, God had delivered them from their enemies and led them into a land of milk and honey. The proof of this covenant was the defeat of the Spaniards and the papists and the material efflorescence that followed. But now there was trouble in Canaan: there were idolaters and schismatics (i.e., the papists and the Arminians) who were worshiping false gods; there were sinners and backsliders who were breaking and ignoring God's laws (e.g., the fornicators and the Sabbath breakers); and there were even some who sought to usurp God's Temple (i.e., erastian magistrates) and attack its chosen guardians (the House of Orange). The actions of these people threatened to provoke God's wrath; and, indeed, signs of this wrath were already evident in various natural disasters and military defeats. The only solution was to renew the covenant and win back God's favor. Otherwise, the republic would be doomed.

One of the first people to articulate this vision of the republic was Willem Teellinck, a Calvinist pastor from Middelburg, the capital city of the province of Zeeland (Engelberts 1973). In a series of influential pamphlets written during the 1620s, Teellinck reminded his readers of the

⁹ Of course, this program was not a "program" in the strict sense; it was not contained in a single document or published by a political party. Rather, it was a set of interrelated goals that were articulated by a group of interconnected authors in various books and pamphlets. The term "program" is thus used in a rather loose way here.

many blessings that God had showered upon the Dutch people (Teellinck 1621a, p. A1), upbraided them for their sins against God (Teellinck 1621b, p. 10), and warned them to change their ways, lest the Netherlands be visited with new disasters and defeats (Teellinck 1624a, 1624b). Had Teellinck been an isolated figure, this would have little historical significance. But he was not. Rather, he was the founding father of the "Further Reformation" (*Nadere Reformatie*), a puritanical movement that sought to put Calvin's teachings into practice and bring his vision of the "Godly commonwealth" (*respublica christiana*) to fruition (Heppe 1879; Brienens et al. 1989). By the middle of the century, the Further Reformation had become a national movement of considerable significance.

The program espoused by Teellinck and his followers during the 1620s was primarily religious in character. The one advanced by his successors, by contrast, was much more political (overviews in Geyl [1947] and Poelhekke [1973]). For them, religious discipline and conformity were not enough. Their demands included (1) a strong army and renewed war with Spain, (2) a strong central state and the limitation of provincial autonomy, and (3) the creation of a hereditary stadtholderate occupied by the House of Orange. The change in the message was accompanied by changes in other areas as well—changes in the tone, in the categories, in the messengers, even in the discourses themselves.

The change in tone was manifest in more explicitly militaristic, chauvinistic, and imperialistic rhetoric. While Teellinck and other writers of the 1620s were certainly not pacifists, neither were they militarists. By contrast, the writers of the post-1648 period firmly embraced war over peace and military strength over fiscal austerity (Knuttel 6758, pp. A2–A3; Knuttel 6852, p. A3). Some even argued that the republic must be prepared to fight an "offensive war" (Knuttel 6852), while others proclaimed that they would "rather be dead than French" (Knuttel 10265, p. 7). Pamphleteers of the post-1648 period also displayed a much more militant and intolerant attitude toward non-Calvinists. Whereas Teellinck had been apt to characterize them as lost sheep, later writers were more likely to describe them as "internal enemies" (*inwendyge vyanden*). Indeed, Teellinck's own son, Maximilliaen, lumped "Evil-Doers, Papists, Remonstrants and other Sectarians" together with one another, calling them "enemies of Religion and State," and accusing them all of betraying "Israel," their "Fatherland" (Teellinck 1650, p. 5). And he was not alone: the idea that the Dutch Catholics were the Trojan horse of a popish plot was a recurring theme in many Calvinist and Orangist pamphlets. The attitude toward followers of the States Party was equally harsh. One particularly virulent pamphlet described the Arminians and Libertines as a "consuming cancer" on the body politic and suggested that the republic "cleanse" itself of such people by means of the scaffold (Knuttel 10499, pp. B2, B4).

For these writers, all “good patriots” were good Calvinists, and all non-Calvinists were potential “traitors.” The hatred that had once been reserved for the Spaniards was now transferred to Libertines, Arminians, papists, and other “enemies of state.” This militarism and intolerance tended to go hand-in-hand with a certain degree of imperialism. In the works of Pietist clerics such as Godefried Udemans and Maximilliaen Teellinck (Brienen 1987), Dutch colonialism was portrayed as a religious obligation. In their view, God had given the Dutch their overseas possessions, and it was the duty of the Netherlands to civilize the “Indians” and bring them God’s word (Udemans 1640; Teellinck 1650). Similar views can be found in the writings of some lay intellectuals (e.g., Sande 1650).

Underlying this shift in tone was a shift in categories. During the revolt and the Quarrels, Calvinists and Orangists often spoke of “the Netherlands” and “the people,” but they rarely used the terms “nation” or “state,” and if they did, they used them in a way that was not entirely modern. Thus, the word “nation” was used mainly in relation to foreign peoples, while the word “state” in the phrase “state of the United Provinces” meant both situation and government.¹⁰

This changed after 1648. More and more pamphleteers began to refer to “the Netherlandish Nation” or simply “our Nation,” and when they discussed the “State of the United Netherlands,” they now meant the central government—and the Netherlandish people (e.g., Knuttel 6740, 6758, 6759).¹¹ What is more, they saw the state and the nation as things to be valued and defended (e.g., Knuttel 6852). It is God who created this nation, argued Maximilliaen Teellinck, “and that is why we must maintain our State for all eternity” (1650, p. 9).

The struggles of the post-1648 period not only transformed the existing Hebraic discourse; they also spawned two entirely new discourses. The first was a discourse regarding the House of Orange. As we have seen, it had become customary, during the days of the Dutch Revolt, to compare the stadtholders to certain Old Testament figures, especially David and Moses. This custom continued into the period under discussion—and beyond it (see, e.g., Tal 1898). During the 1650s, however, some writers began to advance a more grandiose set of claims, arguing that the House of Orange was an instrument of God, that the individual stadtholders possessed special “gifts” that were passed on from father to son, and even that they were divinely “elected” (i.e., saved) by birth (e.g., Teellinck 1650, p. 2; Knuttel 10594, p. 2; Knuttel 10628, p. 11). Of course, claims of this

¹⁰ This multivocality was especially evident in the phrase “state of war” (*Staat van Oorlog*), which meant condition of war and war budget.

¹¹ Indeed, one pamphlet explicitly compares states to peoples, arguing that both have a similar life-course (Knuttel 9893).

sort had long been made about other princes—particularly the kings of France (Kantorowicz 1957; Pange 1949)¹²—but they had never been made about the Dutch stadtholders.

The second new discourse regarded “the People.” Of course, writers of a Hebraicist persuasion had been using the term people for some time in phrases such as “God’s people,” the “Chosen people,” and the “Netherlandish people.” Likewise, writers of a republican persuasion had long argued that the power of princes should be used for the benefit of the people, and, indeed, that the powers of all rulers ultimately derived from the consent of the people (Gelderen 1992). And in popular usage, of course, “people” could be used to mean common people as opposed to rulers or regents (Verwijs, Verdam, and Stoet 1885–1929, no. 9, pp. 842–43). What was unusual about the new usage was not just the fact that it employed capital letters, significant as this was, but the fact that it combined these three meanings: Dutch people, sovereign people, and common people. Thus, during the popular uprisings of 1672, several pamphleteers argued that “the People” had shed their blood to win freedom from the Spanish, that “sovereignty” (*Heerschappye*) therefore belongs to “the People” and not to the regents or town councils, that the latter had gradually usurped the rights and privileges of the former, and that it was now time to restore power to the People by means of popular assemblies (e.g., Knuttel 10564, 10594). It is generally—and incorrectly—believed that the category of the People was not used in this way until the French Revolution.

One might suppose that these two new discourses—royalist and populist—would be used against one another. But this was not the case. It was one of the peculiarities of early modern Dutch politics that monarchy and democracy—demands for a strong stadtholderate and for popular sovereignty—tended to go hand-in-hand with one another. That is because the populists and the Calvinists had a common enemy: the republican regents (Geyl 1947). This alliance between the “little people” and the House of Orange would continue right up through the revolutionary era—and beyond.

The regents of Holland and their supporters responded to this alliance by criticizing the Calvinists’ use of the Hebraic analogy. By comparing the stadtholders to the kings of the Old Testament, the Calvinists were advancing an implicit argument for hereditary monarchy—an argument that was being made more explicitly by Calvinist political theorists at the major Dutch universities (Kossmann 1960, pp. 16–28). Propatrician

¹² It seems quite likely that the new discourse was a French import. During his invasion of the Netherlands in 1627, Louis XIV minted and distributed numerous medals of himself bearing the appellation “most Christian king.” The Dutch would have become familiar with the French cult of monarchy through this and other means and may well have been reacting against it.

writers responded to this argument by pointing out that Yahweh had granted Israel's demand for a king only with great reluctance (De Witt 1654, chap. 14, p. 8): "God did at first mercifully institute no other but a Commonwealth Government, and afterward in Wrath appointed one Sovereign over them." Another pamphleteer put it more baldly: "Republics are more pleasing to God than Monarchies" (Knuttel 5900, p. B). This argument was used by other politicians and political theorists as well (Geyl 1947, pp. 22–23). States writers also drew on the Old Testament to deflect the theocratic pretensions of the Calvinist clergy. Moses, the lawgiver, they pointed out, had been set above Aaron, the chief priest: "God spoke to Moses, Moses to Aaron, and Aaron to the people" (Knuttel 8924, p. 7).¹³

On occasion, states writers would also invoke the Batavian mythology. Thus, one pamphleteer derided the Orangist coup of 1650 as a crime that "the counts and lords [had] never dared essay against the Free Batavians" (Knuttel 6771). But arguments of this sort were surprisingly rare in the works of states writers, probably because the Batavian mythology had been subjected to a withering scholarly attack (e.g., in Schrijver 1646) and was no longer taken very seriously even by its progenitors (Kampinga 1980; Molewijk 1988, pp. 22–24).

As one might expect, the criticisms and theories of the states writers provoked a strong counterresponse from the Calvinists and Orangists; indeed, the two positions developed very much in dialogue with one another. There is no need to review these interchanges in all their intricacies in the present context. What is interesting, from our perspective, is the degree to which the parties tended to appropriate one another's narratives and symbols, albeit to opposed ends. We have already seen that the states writers sometimes spoke in a Hebraic idiom in an attempt to turn the

¹³ This argument was also made in pictorial form in a painting by Ferdinand Bol, which was commissioned for the Town Hall in Amsterdam. The painting showed Moses descending from Mount Sinai, with the Jewish people arrayed at his feet in postures of gratitude and deference (Heppner 1935, pp. 241–45; Blankert 1975, pp. 33–36; Schama 1989, pp. 114–17). The explicit moral is given in a verse by Vondel (1927–40, vol. 8, p. 757), which was affixed to the frame:

Moses of the Hebrews received the Law from God
And with it returned from above to the people
Who greeted and welcomed him with reverence and longing.
A Free State breathes easy, when the people honors the Laws.

The regents, it seems, had recast themselves in the role of Moses. There was also an implicit moral, however, which was made clear by a frieze placed directly below the painting. It shows the Israelites worshipping the Golden Calf during Moses sojourn on Mount Sinai. Without the restraining hand of the elders (Moses), it suggests, the people are misled by the priests (Aaron), and chaos ensues (Blankert 1975, pp. 32–33).

Calvinists own arguments against them. But the reverse was also true. As the century wore on, one finds more and more examples of Calvinists and Orangists invoking the Batavian past, arguing that true Batavians were orthodox and "pro-prince" (*prinsgezind*; e.g., in Knuttel 10569, p. 4; Knuttel 10265), and weaving together Batavian and Hebraic symbols and figures (e.g., in Knuttel 10955; Knuttel 10376). Paradoxically, then, the divergence in the views of the two parties was accompanied by a convergence in their discourses.

Dutch Conclusions: Scope and Content Revisited

The principal argument I have been trying to make in the preceding pages is that the national consciousness of the Dutch in the early modern era was much more modern and much more nationalistic than the modernist position allows. Having made this argument in narrative fashion, I would now like to restate it in analytical terms. At the outset of the article, I reviewed the various criteria the modernists have used to distinguish pre-modern national consciousness from modern nationalism, namely: ideological and categorical *content* and social and political *scope*. Using these criteria, I would now like to assess how modern and how nationalistic early modern Dutch national consciousness really was.

Ideological content.—One of the putative hallmarks of modern nationalism is the view that the world is made up of nations, that these nations are essentially distinct from one another, that one's own nation has a special character or mission, and that it must be sovereign to realize this character or mission. To what degree was this constellation of beliefs present in the early modern Netherlands? As regards the first point, there can be no doubt that a substantial number of Dutch men and women perceived their world, *inter alia*, as a world of nations by the mid-17th century. This much is evident from the confident way in which many pamphleteers spoke of the Spanish nation, the English nation, the French nation, and, of course, the Netherlandish nation. Nor can there be any doubt that many of these writers regarded the various nations as distinct from one another in terms of language, customs, and character. Eventually, some even tried to catalog and explain these differences (e.g., De Hooghe 1706–7). Grotius (1655, p. 7) himself believed that such differences were at the root of the revolt against Spain.

Categorical content.—Another supposed hallmark of modern nationalism lies in the categories of nation, people, and state, and, in particular, in the equation "nation = people (should) = state." Here again the evidence is clear. In Dutch, the word "nation" (*nacie*, *nacioen*, *natie*) becomes a synonym for "people" (*volc*, *volck*, *volk*) in the sense of a group united by ancestry, language, and customs during the 1560s, that is, in the early

years of the Dutch Revolt. This meaning of the world then gradually eclipses earlier meanings during the 17th century (Verwijs et al. 1885–1929, no. 4, pp. 2078–81; Clerck et al. 1882–1998, no. 9, pp. 1586–90). The development of the word “fatherland” (*vaderland*) is similar. During the Middle Ages, it was used to refer to the Land of the Father (i.e., the heavenly kingdom) or to the city of one’s birth. Around the time of the revolt, it starts to be used in a more encompassing way to denote the province of one’s birth or the Netherlands as a whole (Verwijs et al. 1885–1929, no. 8, pp. 1140–41; Clerck et al. 1882–1998, no. 18, pp. 176–80). The development of the word “state” is somewhat slower. It does not acquire its modern meaning in Dutch—a government or the people subject to it—until around the middle of the 17th century (Verwijs et al. 1885–1929, no. 7, pp. 1887–95). Thus, the lexicographical evidence confirms the analysis developed above. In the Netherlands, the categorical elements of modern nationalism were in place by around 1650.

Social scope.—Another feature of modern nationalism that purportedly distinguishes it from premodern national consciousness is its social scope. Whereas premodern national consciousness can be found among certain segments of the elite, it is argued, modern nationalism is a mass phenomenon. In this respect, too, the Dutch case represents a troubling anomaly for the modernist position. For there can be little doubt that the overwhelming majority of 17th-century Netherlanders would have been exposed to nationalist discourses and symbols in one form or another. For those who could read—and that was probably the majority—there were the political pamphlets, which this article has drawn on so heavily. They were a mass medium par excellence—low in price, high in circulation, aimed, as often as not at the “man in the street.” (Revealingly, many of the pamphlets were composed in the form of dialogues between ordinary people—soldiers, shopkeepers, blacksmiths, farmers, and so on.) And for those who could not read, there were various verbal, visual, and ritual media such as sermons, paintings, and public prayers. There cannot have been many Netherlanders who were wholly unfamiliar with the Hebraic and Batavian mythologies, and there were evidently a good many who embraced them as well.

Political scope.—A fourth and final criteria modernists use to distinguish premodern national consciousness from modern nationalism is political mobilization. Before the modern era, it is argued, national consciousness never led to nationalist politics. Here, too, the modernists appear to have overstated their case. For it would be hard to deny that the political demands of the Calvinists and Orangists did not contain nationalist elements: the persecution of internal enemies, the maintenance of a strong military, the reunification of North and South, the defense of the overseas empire, and so on. Nor was the States Party immune to nationalistic fer-

vor. For them, the distinguishing features of the Netherlanders were commercial prowess and political virtue. Indeed, one could see the debates between the two parties as debates over the true identity of the Netherlandish nation. By the standards of the modernists themselves, then, the national consciousness of the early modern Netherlanders looks suspiciously like modern nationalism.

EXODUS REDUX: MEDIEVAL, EARLY MODERN, AND MODERN NATIONALISMS COMPARED

By itself, the Dutch case does not really suffice to undermine the modernist position, since it could very well be that the Netherlands were just an anomaly, one of those "odd predecessors" to modern nationalism of which Hobsbawm (1990, p. 3) speaks. If this were so, the modernist argument would be somewhat weakened, but it could still be defended. But it is not so, and, ironically enough, it is in England, Hobsbawm's adopted homeland, that we find one of the most clear-cut cases of early modern nationalism.

"Many Are Chosen . . .": The Mosaic Moment in North Atlantic Nationalism

One of the first historians to identify a link between Protestantism and nationalism in early modern England was William Haller. In his well-known study of John Foxe's *Book of Martyrs* (1832), an account of the "lives and sufferings" of the Christian martyrs from antiquity to the present, Haller (1963; see also Firth 1979) showed that Foxe accorded a special place to England in the unfolding of the divine plan: for Foxe, England was an "elect nation," comparable in significance to the Israel of the Old Testament. It must have been an appealing view. For as Haller shows, the *Book of Martyrs* went through four editions between 1554 and 1570 and became a fixture in all parish churches after 1571: from that year on, every church was required to purchase and display a copy, by order of convocation. And Foxe's work was not the only place where the English laity would have encountered the Hebraic idiom. As recent scholarship has shown, Hebraic imagery abounded in early modern England. It could be found in "godly ballads," political pamphlets, public rituals, Protestant preaching, and popular prints, to name only the best-documented media (see e.g., Colley 1992; Watt 1991; Hill 1993; Cressey 1989; Bradshaw 1996; Collinson 1997). Given the plethora of sources in which it was invoked, it seems reasonable to assume that most English Protestants would have encountered the Hebraic idiom sometime during the Elizabethan period.

Up until the late 16th century, the idea of national election seems to

have been a source of unity among English Protestants. But by the early 17th century, it had become a source of division (Collinson 1988, p. 23). For the "hotter sort of Protestants," the Puritans, the special favors God had granted to the English nation (e.g., affluence and liberty) implied a special burden upon the English people (i.e., a national covenant). If the English should shirk this burden, they would be harshly punished, for "God punisheth national sins with national punishments" (quoted in Hill [1993, p. 285]). Among the cooler sort of Protestants, by contrast, Hebraic rhetoric was decidedly *outré*. The Protestant jurist, Richard Hooker, regarded comparisons between England and Israel as self-aggrandizing and ill-informed (Hill 1993, p. 267). And Arminian churchmen, such as William Laud, disputed the idea of collective election upon which such analogies rested. By the late 16th century, then, English Protestantism was beginning to split apart into two opposing movements, Puritanism and Anglicanism, each with its own distinctive view of the English nation (Tyacke 1986).

Following the accession of Charles I in 1625, the struggle between these two movements within the Church rapidly evolved into a struggle between two "religious parties" in Parliament, a parliamentary party associated with the Puritans and a loyalist party associated with the Anglicans (Russell 1979, 1990; Hunt 1983). And with the onset of the war with Scotland in 1640, the struggle between these two parties exploded into civil war. Historians used to refer to these events as "the Puritan Revolution," but they could also be seen as a nationalist coup. Because what the Roundheads were fighting for, and what they sought to achieve, was the renewal of the national covenant, which the English people had entered into with God; and what they were fighting against, and what they overthrew, was a pro-Spanish and crypto-Catholic court—or so they believed. Similarly, one might also view the English Republic as a "nationalizing state" (Brubaker 1996). From this perspective, the efforts of the new regime to impose godly morality on the English people were an effort to enforce the terms of the national covenant, and Cromwell's campaigns in Ireland and the West Indies were not imperialistic ventures so much as religious crusades—wars against the infidels.

The Restoration of 1660 brought an end to the Cromwellian era and its Puritanical policies. But it did not bring an end to the Hebraic myth. On the contrary, Charles II was careful to portray himself in Hebraic terms. "By ways and means no less miraculous than those by which He did heretofore preserve and restore his own chosen people," he proclaimed, "it hath pleased [God] . . . to restore Us and Our good subjects to each other" (Hill 1993, p. 269; see also Pincus 1996). And the revolution of 1689 brought a full-fledged Hebraic revival. William III and his supporters consistently used Hebraic rhetoric to legitimate their usurpation of James

II and his crypto-Catholic court, and they explicitly framed their support for a national "reformation of manners" as an attempt to restore the "national covenant" (Claydon 1996). In learned circles, the Hebraic idiom gradually fell out of favor following the Restoration (Zwicker 1988). But it lived on in popular discourse. As late as 1756, a Protestant preacher could still ask: "Do we not succeed the Jews? . . . Are not we a chosen generation, a peculiar people, as they were?" (Collinson 1988, p. 7).

As in the Netherlands, there were also other strands of nationalist discourse. One was classical. It posited an essential continuity between ancient Britannia and modern England in the realms of culture, law, politics, and religion (Pocock 1987; Helgerson 1992; Kunst 1994; see also Parsons 1929). This discourse might be called the Britannian myth. There was also a second variant of this myth. During the 17th century, republican authors often portrayed England as the new Rome, the heir apparent of classical civilization (Fink 1945; Worden 1982; see also Pocock 1975; Smuts 1993; Salmon 1991). The second discourse was dynastic. In England, as in France, there was a tradition of "sacred monarchy" whose history stretched back into the early Middle Ages (Kantorowicz 1957; Bloch 1973). Following the Reformation, however, this tradition was strengthened and transformed: strengthened by the subjugation of the church to the crown and transformed by the ascension of a woman to the throne—the "Virgin Queen," Elizabeth I, whose birthday became the subject of a veritable cult of personality, and whose ascension was celebrated like a religious holiday (Cressey 1989, pp. 50–64; Hackett 1995). The third, finally, was democratic in character. It was also a product of the English Civil War and portrayed the English as a sovereign people who occupied a special place within the community of nations (Zagorin 1966; Wootton 1986, pp. 21–77; Greenfeld 1992).

In many ways, Dutch and English nationalism were remarkably similar during this period.¹⁴ Initially, both were articulated in a Hebraic idiom. Both were closely connected to Calvinism. Both gave rise to controversy and counternarratives. And both eventually became entwined with other discourses—classical, monarchical, and democratic.

Nor do these parallels end with England. Similar points could be made about Scotland and America, too. For the Hebraic discourse of "cho-

¹⁴ This is not to say that they were identical. There were several important differences. For example, the word "nation" appears much more frequently in English writings. Dutch authors were more apt to use the word "people" (*volk*). Also, the monarchical discourse was more important in England—as was the monarchy, itself. In England, finally, the Hebraic discourse was the official discourse of the regime, at least under Cromwell and William III. In the Netherlands, by contrast, it was mainly an oppositional discourse. Thus, there were differences in both the content and scope of nationalism in these two polities.

senness," "deliverance," and "covenant" can also be found in the rhetoric surrounding the Scottish uprisings and the American Revolution. Indeed, the Scottish Calvinists were probably the first to invoke the covenant idea, and it is from them that the English most likely derived it (Oestreich 1969; Cowan 1976). From there, it was transported to the soil of the Americas (O'Brien 1988; Moorhead 1994; Cherry 1971; Hudson 1970). And Dutch Hebraicism had its offspring, too: the Hebraic idiom figured prominently in the nationalist rhetoric of the Boers (Toit 1994).

In his widely read and highly influential study of republican thought in North Atlantic Europe, J. G. A. Pocock (1975) speaks of a "Machiavellian moment," by which he means a moment in historical time and political debate in which the Machiavellian equation of citizen, *civitas* and *virtù*, emerges with sudden clarity and effect. Perhaps one could also speak of a "Mosaic moment" in the history of North Atlantic nationalism and the Anglo-Dutch archipelago.

"... But Few Are Called": Calvinism and National Chosenness?

Based on the foregoing examples, it would be tempting to conclude that the Hebraic mythology was the form in which Western nationalism first crystallized, that this crystallization was distinctly early modern, and that it was precipitated by Calvinism—which was the most obvious thing the North Atlantic polities had in common. Alas, however, these conclusions prove too simplistic. There are several reasons for this.

As it turns out, the roots of Hebraic discourse extend back well beyond the early modern period. Speaking before Parliament in 1377, for example, Richard, the chancellor of England, roundly proclaimed that "God would have never honored this land in the same way as he did Israel through great victories over their enemies, if it were not that He had chosen it as his heritage" (McKenna 1982, p. 31; see also Turville-Petre 1996; Hastings 1997). Similar claims were also made by the French at this time; indeed, Richard's statement was a direct response to these claims. Medieval examples of Hebraic discourse could also be cited for Portugal (Marcu 1976, p. 14),¹⁵ Italy, and Germany (O'Brien 1988, pp. 12–22) and could probably be found for a good many other countries as well. In all likelihood, the Hebraic idiom was well-nigh universal in Medieval Europe.

This raises an important question: Why was it that the Hebraic mythology became a part of early modern political discourse in some countries,

¹⁵ We will soon have a good deal more information on this subject, since the pervasiveness of Hebraic imagery in early modern Portugal also forms the subject of a dissertation in progress by Eric Olsen, a graduate student in history at the University of Pennsylvania.

but not in others? Based on the foregoing analysis, one possible answer would seem to be Calvinism. After all, that was one of the most obvious things the North Atlantic nations had in common. Unfortunately, there were some important exceptions to this rule. There was at least one Catholic country (Portugal) and one Lutheran country (Hessia) that exhibited Hebraic forms of nationalism, and there was at least one (partly) Calvinist country (Brandenburg-Prussia) that did not. This suggests another possible explanation: the experience of popular revolution. That is one thing that distinguished Portugal and Hessia from Prussia and linked them to the North Atlantic nations. But further research will be necessary before this question can be answered conclusively.¹⁶

France: How National Was Revolutionary Nationalism?

Thus far, I have tried to show that there are at least some examples of early modern national consciousness that fulfill the definitional criteria of modern nationalism laid out by the modernists, and that early modern national consciousness had medieval roots. To put it more pointedly, I have tried to show that the modernist position rests on a distorted—one might even say caricatured—picture of early modern political culture, one which *understates* the degree of nationalism during this period. In this section, I reverse the procedure: I try to show that the modernist position also rests on a distorted picture of the French Revolution, one which *overstates* the novelty of nationalist discourse and the scope of nationalist mobilization that occurred during the years after 1789.

If the evidence I have presented above is correct, then it is no longer possible to argue that there were no truly nationalist discourses and no fully nationalist politics until the French Revolution. As we have seen, the Dutch Revolt and the English Civil War both gave rise to discourses and movements that fulfilled the meaning and scope criteria set forth by the modernists, and more detailed analysis of the Scottish Covenanters and the American Revolution would probably yield similar conclusions. Nonetheless, it could still be the case that the modernist argument is correct in another sense: it could be that the nationalistic content of revolutionary discourse was so much purer, and the scope of nationalist mobilization so much broader, that we must still draw a sharp line between revolutionary and prerevolutionary nationalisms.¹⁷

¹⁶ At the moment, very little is known about the relationship between religion, revolution, and national identity in the Lutheran and Catholic contexts, so it is quite possible that these conclusions simply reflect our ignorance of the historical record. It may even be that the question itself is wrong: perhaps the Hebraic discourse was more widespread in early modern Europe than we presently believe.

¹⁷ Indeed, this is precisely the argument that has been advanced by John Breuilly. "In the political rhetoric of the [early modern] period," he claims, "the idea of the nation

Let us consider these hypotheses in greater detail, beginning with the thesis of discursive purity. At first glance, it would seem to have much to recommend it. After all, there can be no doubt that early modern nationalism was intertwined with religious and monarchical discourses in a way that revolutionary nationalism was not. On second glance, however, the thesis of discursive purity appears less plausible. For revolutionary nationalism was also intertwined with other discourses: the discourses of the *philosophes* regarding civic virtue, social contract, and natural right (Blum 1986), the discourses of the schools regarding classical antiquity, and especially republican Rome (Parker 1965; Schama 1989), and perhaps still others.

What about the scope of nationalist mobilization? Here, too, the evidence is not so clear-cut as the modernists claim. Revolutionary nationalism met with a relatively muted response from the French peasantry (Jones 1988), especially after the "abolition of feudalism" in 1790 (Markoff 1996), and it sparked explicitly antinational rebellions in many parts of the South and the West, the *Vendée* and *chouannerie* being only the two best-known examples (Godechot 1961; Tilly 1964; Dupuy 1988). Opposition to the French Revolution was also manifest in high rates of desertion and "draft dodging" (*insoumission*; Forrest 1989, 1990) and in the unwillingness of many clergymen to swear their loyalty to the nation (Tackett 1986), often with the support of their parishioners (Desan 1990). Of course, one could cite similar examples of opposition for the Dutch Revolt and the English Civil War. But that is precisely the point: it is not immediately evident that the social and political scope of nationalist mobilization within revolutionary France was really that much greater than in 17th-century England or 16th-century Holland.

Does that mean that there were *no* differences between early modern and revolutionary nationalism? Certainly not. There were differences, first of all, in discursive content. For if the revolutionary discourse of the nation was no purer than its early modern predecessors, it was considerably more secular: the early modern nation was based (mainly) on a divine covenant; by contrast, the revolutionary nation was premised (mainly) on a social contract.¹⁸ The *Leitfaden* of early modern nationalism was the

... was subordinated to religious and monarchical principles" (Breuilly 1982, p. 45). Moreover, the Revolution led to a "more continuous and sustained" level of "mass political activity . . . than had been possible in early modern Europe" (Breuilly 1982, p. 49). Revolutionary nationalism, he contends, was distinguished by the purity of its discourse and the intensity of its mobilization (similar arguments may be found in Gellner [1983, pp. 40–41], Hroch [1985], Smith [1986, p. 11], and Hobsbawm [1990, pp. 14–18]; for a critique, see also Smith [1991]).

¹⁸ However, even this difference was largely a matter of degree. As we have seen, early modern nationalism also had a secular strand: the classical mythology. And as Dale Van Kley (1996) has recently shown, French revolutionary ideology had a reli-

Hebraic discourse; that of revolutionary nationalism, the democratic discourse.¹⁹ There were also differences in geographical scope. The French Revolution was a pan-European, even world-historical, event (Palmer 1959–64, p. 2; Godechot 1983) in a way that the English Civil War and the Dutch Revolt were not. Thus, the French Revolution helped to usher in a new *form* of nationalism, and it helped to propagate and popularize nationalist discourse and practices. Whether or not it helped to usher in a new *age* of nationalism, however, is an open question. It seems possible, even probable, that nationalist discourse was more predominant and nationalist mobilization more frequent after the Revolution than before. Only a comparative analysis of political culture and collective action in the modern and premodern eras could resolve this question. While I cannot possibly undertake such an analysis in the context of this article, I would like to suggest how one might go about it. It is to that task that I now turn.

SUMMARY AND CONCLUSION: OUTLINE OF A POSTMODERNIST THEORY OF NATIONALISM

Modernists draw a sharp distinction between premodern national consciousness and modern nationalism. While protonationalistic forms of thought and action may be found among elite groupings during the premodern era, they argue, fully nationalistic ideologies and movements with a mass following do not appear until the time of the French Revolution. Thus, they conclude, premodern national consciousness and modern nationalism differ substantially in their ideological and categorical content and their social and political scope.

The primary purpose of this article has been to challenge the modernist view, by showing that the differences between premodern national consciousness and modern nationalism were not so great as the modernists have supposed, at least not in the Calvinist countries of the North Atlantic. In particular, I have tried to show that the Renaissance and Reformation, and the revolutions that followed in their wake, gave rise to popular discourses and political movements, which must be categorized as nationalistic by the modernists own definitions and criteria. At the same time,

gious strand that stretched back to Calvin (on this, see also Bickart [1932]). There was also a religious element in revolutionary practice. The attempts of the revolutionaries to establish a new, civic religion are well known (Aulard 1892; Mathiez 1904). More recent research has shown that the cult of revolutionary "saints" and the forms of revolutionary festivals were also strongly influenced by Catholic models (Soboul 1957; Ozouf 1988).

¹⁹ The connecting thread, as some readers may have recognized, was the classical discourse.

I have also tried to show, albeit in less detail, that early modern nationalism had medieval roots, and that French revolutionary nationalism was not quite so revolutionary as the modernists suggest, that is, that it was not really that much more national or nationalistic than the Hebraic and classical nationalisms that we observe in the Netherlands and England during the 17th century.

Once we abandon the modernist thesis, as I believe we must, we are confronted with a number of thorny questions. For example, how are we to distinguish between different forms of nationalism—modern versus early modern, Dutch versus English, and so on? Indeed, how are we to distinguish nationalist discourses and practices from nonnationalist ones? In other words, if we can no longer use the French Revolution as our paradigm, then what should our point of reference be? In what remains of this article, I would like to present some analytical strategies and conceptual tools that might be used to construct a postmodernist theory of nationalism.

First, instead of drawing sharp distinctions between protonationalism and fully developed nationalism, I would suggest that we focus on variations in the intensity and scope of nationalist mobilization—thereby expunging the hidden remnants of modernization theory from the theory of nationalism. As regards the intensity of nationalist mobilization, one might distinguish between (1) *discourses* that invoke “the nation” or its kindred categories (people, fatherland, state) or discuss the differences between, or development of, various nations; (2) *movements* that see “the nation”—its preservation, purification, expansion, and so on—as one of their central goals; (3) *parties* that have some degree of central organization and seek to capture or influence the state in the name of “the nation”; and (4) *regimes* that control the central state apparatus and strike out violently against internal and external “enemies of the nation.” As regards the scope of nationalist mobilization, one could distinguish between discourses, movements, parties, and regimes whose carriers or followers include members of (1) intellectual elites (clergy and lay); (2) political or social elites (nobles, patricians, etc.); (3) “the middling sort” (professionals, merchants, craftsmen, etc.); and (4) the common people (peasants, artisans, laborers, etc.). In this schema, the lowest level of nationalist mobilization would be indicated by the existence of a nationalist discourse that was limited to an intellectual elite, while the highest level would be indicated by the existence of a nationalist regime with a mass base. Renaissance Italy would be an example of the former, Nazi Germany an example of the latter. This approach could also be used to code and categorize the various cases treated in this article.

Second, instead of looking for some essential characteristic or roster of characteristics that distinguish real nationalism from pseudo nationalism,

I would suggest that we try to identify the various "threads" of which "the fabric of [modern] nationalist discourse" is composed, recognizing that each of these threads has its own particular history (Calhoun 1997, p. 9), and thereby dispensing with the assumption that nationalism has any natural or inherent unity. In order to facilitate the task of identifying and classifying nationalisms, one could distinguish between nationalist discourse as a whole (the "fabric"), nationalist discourses specific to a certain time or place (the "threads"), the various narratives of which these discourses are composed (the "fibers"), and the categories that hold the narratives together (the "raw materials"). The historical construction of nationalist discourse, to continue the metaphor, could then be conceptualized as a continuous process of harvesting the fruits of the national category, straightening the categorical fibers into linear narratives, spinning these narratives together into longer threads, and (sometimes) weaving these threads together into a larger cloth. From this perspective, the unity and coherence of nationalist discourse, insofar as it exists at all, is a function, not of some essential meaning, a particular set of claims (or principles or doctrines), but of its surface characteristics, the particular set of *categories* (in the West: nation, people, fatherland, state, etc.) around which it is constructed. What gives nationalist discourse its unity and continuity in this schema is not the signified (the meaning) but the signifier (the category). A nationalist discourse, in this schema, is simply a discourse that invokes "the nation" or its kindred categories, and what distinguishes nationalist discourses from one another is the narratives that they employ (the "fibers") and the specific way in which they spin them together (into "threads"). The scholar's job is to describe and explain the results of this process, to show how and why a particular fabric, thread, or fiber looks the way that it does.

This job can be broken down into two discrete but complementary tasks. The first is analytical and historical. It involves identifying the individual narrative fibers that serve, or have served, as the raw materials for particular discursive threads, and the process by which these threads have, or have not, been woven together into a coherent discursive fabric. This task is analytical insofar as it requires one to disentangle a particular narrative fiber from the knotty web of cultural and political debate, in which discourses about the nation are inevitably intertwined with other discourses (discourses about religion, history, democracy, and so on). And it is historical insofar as it forces one to focus on continuity and development, on the unfolding of a particular narrative fiber, the spinning of a particular national thread, or the weaving of the entire nationalist fabric. The second task that confronts the analyst of nationalism is critical and genealogical. It involves testing the strength of the various threads and examining the constituent fibers for ruptures and tears. It is critical insofar

as it requires one to examine just how much historical weight a particular fiber or thread can bear, that is, the degree to which the narrative threads are held together by flimsy historical fictions. And it is genealogical insofar as it forces one to look carefully for weak spots and discontinuities in the fibers, threads, and materials, that is, for changes in the content and meaning of nationalist categories, narratives, and discourses. As Foucault has rightly remarked, genealogical "knowledge is not made for understanding; it is made for cutting" (1977, p. 154).

In closing, let me contrast the modernist and postmodernist positions by showing how each addresses the three questions that any theory of nationalism must address: What is nationalism? When has it happened? And why does it happen?

Modernists generally answer the what question in essentialist terms. They advance definitions that focus on some feature or features of nationalism that they regard as essential, usually having to do with its content or scope, and then use these definitions to distinguish real nationalism from pseudo nationalism. I regard this approach as hopelessly misguided. What is needed, I would argue, is not a deep definition, but a superficial one, a definition that focuses on the signifiers, rather than on the signified. In this spirit, one might define nationalism as *any set of discourses or practices that invoke "the nation" or equivalent categories*. This definition may strike some modernists as overly vague and theoretical. But I would argue that a vague and theoretical definition is all that is possible given the diversity and heterogeneity of nationalist discourses and practices, and that what we need in order to sort these discourses and practices out is not general definitions, but specific classifications.²⁰ In developing these classifications, we should also be wary of the dichotomous schemas propounded by the modernists. What is needed is not a simplistic dichotomy that paints the history of nationalism in black-and-white, before-and-after terms, but a complex framework that can capture the variety and history of nationalism in its many hues of gray.

As regards the when question, modernists usually locate the advent of nationalism in a specific place and time, to wit: France in 1789. And not without reason: as we have seen, a number of important shifts in nationalist discourse and practice did occur in France at this time. But one cannot localize the origins of nationalism, or even of modern nationalism, in this way. For the discourses and practices of which modern nationalism is constructed have long histories of their own, histories that extend back to classical Rome and ancient Israel, and beyond that into the darkness of prehistory. Some modernists have recognized this, of course (e.g., Kohn

²⁰ The problem with the essentialist approach is that it conflates the tasks of (theoretical) definition and (historical) classification.

1967; Smith 1986; Armstrong 1982). But even they have tried to draw sharp lines between the modern and premodern eras. I believe this approach is mistaken. In my view, nationalism does not have an origin or a single history; it has a genealogy, a ruptured and fragmented history whose only unity lies in the national category itself. Nationalism is not something that happened at a particular place and time; it is something that happens in many places and times, and in many different ways. The task of the genealogist is to describe these happenings and to trace out the connections between them. For the modernist, then, the *when* question has one answer; for the genealogist, by contrast, it has many.

REFERENCES

- Anderson, Benedict. 1992. *Imagined Communities*, 2d ed. London: Verso.
- Armstrong, John A. 1982. *Nations before Nationalism*. Chapel Hill: University of North Carolina Press.
- Aulard, F. A. 1892. *Le culte de la raison et le culte de l'Être Suprême (1793–1794)*. Paris: Alcan.
- Balibar, Etienne. 1991. "The Nation Form: History and Ideology." Pp. 86–106 in *Race, Nation, Class: Ambiguous Identities*, edited by Etienne Balibar and Immanuel Wallerstein. New York: Verso.
- Beaune, Colette. 1985. *Naissance de la nation France*. Paris: Gallimard.
- Berding, Helmut, ed. 1994. *Nationales Bewusstsein und kollektive Identität: Studien zur Entwicklung des kollektiven Bewusstseins in der Neuzeit*, vol. 2. Frankfurt am Main: Suhrkamp.
- Beumann, Helmut, ed. 1983. *Beiträge zur Bildung der Französischen Nation im Früh- und Hochmittelalter*. Sigmaringen: J. Thorbecke.
- Beumann, Helmut, and Werner Schröder, eds. 1978. *Aspekte der Nationenbildung Im Mittelalter: Ergebnisse d. Marburger Rundgespräche 1972–1975*. Sigmaringen: J. Thorbecke.
- Bickart, Roger. 1932. *Les parlements et la notion de souveraineté au XVIIIe siècle*. Paris: F. Alcan.
- Bijl, M. Van der. 1981. *Idee en interest: voorgeschiedenis, verloop en achtergronden van de politieke twisten in Zeeland en vooral in Middelburg tussen 1702 en 1715*. Groningen: WoltersNoordhoff.
- Blankert, Albert. 1975. *Kunst Als Regeringszaak in Amsterdam in de 17e eeuw: Rondom schilderijen van Ferdinand Bol*. Lochem: De Tijdstroom.
- Bloch, Marc. 1973. *The Royal Touch: Sacred Monarchy and Scrofula in England and France*. London: Routledge & Kegan Paul.
- Blum, Carol. 1986. *Rousseau and the Republic of Virtue: The Language of Politics in the French Revolution*. Ithaca, N.Y.: Cornell University Press.
- Bradshaw, Brendan, and Peter Roberts, eds. 1998. *British Consciousness and Identity: The Making of Britain, 1533–1707*. Cambridge: Cambridge University Press.
- Bradshaw, Christopher. 1996. "David or Josiah? Old Testament Kings as Exemplars in Edwardian Religious Polemic." Pp. 77–90 in *Protestant History and Identity in Sixteenth-Century Europe*, vol. 2. Edited by Bruce Gordon. Aldershot: Scolar Press.
- Brass, Paul R. 1991. *Ethnicity and Nationalism: Theory and Comparison*. New Delhi: Sage.
- Breuille, John. 1982. *Nationalism and the State*. Chicago: University of Chicago Press.
- Brienen, T., et al., eds. 1987. *Figuren en thema's van de Nadere Reformatie*. Kampen: De Groot Goudrian.

- . 1989. *De Nadere Reformatie en het gereformeerd piëtisme*. The Hague: Boekcentrum.
- Brubaker, Rogers. 1996. *Nationalism Reframed*. Cambridge: Cambridge University Press.
- Calhoun, Craig. 1997. *Nationalism*. Minneapolis: University of Minnesota Press.
- Cherry, Conrad. 1971. *God's New Israel: Religious Interpretations of American Destiny*. Englewood Cliff, N.J.: Prentice-Hall.
- Claydon, Tony. 1996. *William III and the Godly Revolution*. Cambridge: Cambridge University Press.
- Clerck, W De, et al. 1882–1998. *Woordenboek der nederlandsche taal*. Leiden: Nijhoff.
- Colley, Linda. 1992. *Britons: Forging the Nation 1707–1837*. New Haven, Conn.: Yale University Press.
- Collinson, Patrick. 1988. *The Birth Pangs of Protestant England: Religious and Cultural Change in the Sixteenth and Seventeenth Centuries*. New York: St. Martin's Press.
- . 1997. "Biblical Rhetoric: The English Nation and National Sentiment in the Prophetic Mode." In *Religion and Culture in Renaissance England*, edited by Claire McEachern and Debra Shuger. Cambridge: Cambridge University Press.
- Cowan, Ian B. 1976. *The Scottish Covenanters, 1660–1688*. London: V. Gollancz.
- Cressey, David. 1989. *Bonfires and Bells: National Memory and the Protestant Calendar in Elizabethan and Stuart England*. Berkeley and Los Angeles: University of California Press.
- Dann, Otto, ed. 1986. *Nationalismus in vorindustrieller Zeit*. Munich: R. Oldenbourg.
- Desan, Suzanne. 1990. *Reclaiming the Sacred: Lay Religion and Popular Politics in Revolutionary France*. Ithaca, N.Y.: Cornell University Press.
- Deursen, A. Th. Van. 1991. *Bavianen en slijkgeuzen: kerk en kerkvolk in den tijd van Maurits en Oldenbarnevelt*. Franeker: Van Wijnen.
- Deutsch, Karl W. 1966. *Nationalism and Social Communication: An Inquiry into the Foundations of Nationality*, 2d ed. Cambridge, Mass.: MIT Press.
- De Witt, Johan. 1654. *Deductie, Ofte Declaratie Van De Staten Van Hollandt Ende West-Vrieslandt*. The Hague: Hillebrant Jacobsz. Van Wouw.
- Dupuy, Roger. 1988. *De la révolution a la chouannerie: paysans en Bretagne, 1788–1794*. Paris: Flammarion.
- Dutch Pamphlets. 1486–1750. Section 1, the Collection in the Royal Library, The Hague, 1979–82. Zug: IDC
- Duyse, Fl. Van. 1965. *Het oude nederlandsche lied*. Hilversum: Frits A. M. Knuff.
- Ecclesiastical History Society. 1982. *Religion and National Identity: Papers Read at the Nineteenth Summer Meeting and the Twentieth Winter Meeting of the Ecclesiastical History Society*, edited by Stuart Mews. Vol. 18 in *Studies in Church History*. Oxford: Basil Blackwell.
- Ehlers, Joachim, ed. 1989. *Ansätze und Diskontinuität deutscher Nationsbildung im Mittelalter*. Sigmaringen: J. Thorbecke.
- Eichenberger, Thomas, ed. 1991. *Patria: Studien zur Bedeutung des Wortes im Mittelalter (6. bis 12. Jahrhundert)*. Sigmaringen: J. Thorbecke.
- Engelberts, W. J. M. 1973. *Willem Teellinck*. Amsterdam: Ton Bolland.
- Fink, Zera Silver. 1945. *The Classical Republicans: An Essay in the Recovery of a Pattern of Thought in Seventeenth Century England*. Evanston, Ill.: Northwestern University Press.
- Firth, Katherine R. 1979. *The Apocalyptic Tradition in Reformation Britain, 1530–1645*. Oxford: Oxford University Press.
- Forde, Simon, Lesley Johnson, and Alan V. Murray, eds. 1995. *Concepts of National Identity in the Middle Ages*. Leeds: University of Leeds Press.

- Forrest, Alan I. 1989. *Conscripts and Deserters: The Army and French Society during the Revolution and Empire*. New York: Oxford University Press.
- . 1990. *The Soldiers of the French Revolution*. Durham, N.C.: Duke University Press.
- Foucault, Michel. 1977. *Language, Counter-Memory and Practice: Selected Essays and Interviews*, edited by Donald Bouchard. Ithaca, N.Y.: Cornell University Press.
- Foxe, John. 1832. *Book of Martyrs*. Cincinnati, Ohio: Roff and Young.
- Gelderen, Martin van. 1992. *The Political Thought of the Dutch Revolt, 1555–1590*. Cambridge: Cambridge University Press.
- . 1993. *The Dutch Revolt*. Cambridge: Cambridge University Press.
- Gellner, Ernst. 1983. *Nations and Nationalism*. Ithaca, N.Y.: Cornell University Press.
- Geyl, Pieter. 1934. "Godsdienst en nationaliteitsgevoel in de Nederlanden ten tijde van Frederik Hendrik." *Bladeren voor vaderlandsche geschiedenis en oudheidkunde* 7 (4): 29–41.
- . 1947. "Het stadhouderschap in de partij-literatuur onder De Witt." *Mededeelingen der Koninklijke Nederlandsche Akademie van Wetenschappen, Afd. Letterkunde* 10 (2): 17–68.
- . (1930) 1975. "Opkomst en verval van het noord-nederlands nationaliteitsbesef." Pp. 17–33 in *Vaderlands verleden in veelvoud: 31 opstellen over de nederlandse geschiedenis na 1500*, edited by G. A. M. Beekelaar, et al. The Hague: Nijhoff.
- Giesen, Bernhard, ed. 1991. *Nationale und kulturelle Identität: Studien zur Entwicklung des kollektiven Bewusstseins in der Neuzeit*. Frankfurt am Main: Suhrkamp.
- Godechot, Jacques L. 1961. *La contre-révolution, doctrine et action, 1789–1804*. Paris: Presses universitaires de France.
- . 1983. *La Grande Nation: L'Expansion révolutionnaire de la France dans le monde de 1789 À 1799*, 2d ed. Paris: Aubier Montaigne.
- Greenfeld, Liah. 1992. *Nationalism: Five Roads to Modernity*. Cambridge, Mass.: Harvard University Press.
- Groenhuis, G. 1977. *De Predikanten: De sociale positie van de gereformeerde predikanten in de Republiek der Verenigde Nederlanden vóór [Ca.] 1700*. Groningen: Wolters-Noordhoff.
- Groenveld, S. 1967. *De prins voor Amsterdam. Reacties uit pamfletten op de aanslag van 1650*. Bussum: Fibula-Van Dishoeck.
- . 1990. *Evidente factiën in den staet*. Hilversum: Verloren.
- Groenveld, S., et al. 1991. *De tachtigjarige oorlog*, 3d ed. Zutphen: De Walburg Pers.
- Grosby, Steven. 1991. "Religion and Nationality in Antiquity: The Worship of Yahweh and Ancient Israel." *Archives européennes de sociologie* 32:229–65.
- Grotius, Hugo. 1655. *Of the Law of Warre and Peace*. London: T. Warren, for William Lee.
- Hacket, Helen. 1995. *Virgin Mother, Maiden Queen: Elizabeth I and the Cult of the Virgin Mary*. New York: St. Martin's Press.
- Haller, William. 1963. *Foxe's Book of Martyrs and the Elect Nation*. London: J. Cape.
- Hastings, Adrian. 1997. *The Construction of Nationhood: Ethnicity, Religion and Nationalism*. Cambridge: Cambridge University Press.
- Hayes, Carlton J. H. 1931. *The Historical Evolution of Modern Nationalism*. New York: Macmillan.
- Heissenbüttel, Kurt. 1920. *Die Bedeutung der Bezeichnungen für 'Volk' und 'Nation' bei den Geschichtsschreibern des 10. bis 13. Jahrhunderts*. Göttingen: Dietrichsche Universitäts-Buchdruckerei.
- Helgerson, Richard. 1992. *Forms of Nationhood: The Elizabethan Writing of England*. Chicago: University of Chicago Press.

- Heppner, A. 1935. "Mozes zeigt die Gesetztafeln bei Rembrandt und Boll." *Oud-Holland* 52:241-45.
- Hill, Christopher. 1993. *The English Bible and the Seventeenth Century Revolution*. London: Allen Lane.
- Hobsbawm, Eric. 1990. *Nations and Nationalism Since 1780*, 2d ed. Cambridge: Cambridge University Press.
- Hooft, P. C. 1677. *Nederlandsche Historien, Seedert de Ooverdracht der Heerschappye Van Kaizar Kaarel Den Vyfden Op Kooning Philips Zynen Zoon, Tot de Doodt Des Prinsen Van Oranje*, 3d ed. Amsterdam: Johan van Someren, Abraham Wolfgangk, Henrik en Dirk Boom.
- . 1899-1900. *Verzamelde Gedichten*, 2d rev. ed. Edited by F. A. Stoett and P. Leendertz. Amsterdam: P. N. Kampden & Zoon.
- . (1626) 1954. *Baeto*. Zwolle: W. E. J. Tjeenk Willink.
- Hooghe, Romeyn De. 1706-7. *Spiegel Van Staat Des Vereenigde Nederlands*. Amsterdam: Ian ten Hoorn.
- Hroch, Miroslav. 1985. *Social Preconditions of National Revival in Europe*. Cambridge: Cambridge University Press.
- Hudson, Winthrop S. 1970. *Nationalism and Religion in American: Concepts of American Identity and Mission*. New York: Harper & Row.
- Huisman, C. 1983. *Neerlands Israël: het natiebesef der traditioneel-gereformeerden in de achttiende eeuw*. Dordrecht: J. P. van den Tol.
- Hunt, William. 1983. *The Puritan Moment: The Coming of Revolution in an English County*. Cambridge, Mass.: Harvard University Press.
- Israel, Jonathan. 1995. *The Dutch Republic: Its Rise, Greatness and Fall, 1477-1806*. Oxford: Oxford University Press.
- Jones, P. M. 1988. *The Peasantry in the French Revolution*. Cambridge: Cambridge University Press.
- Kahl, Hans-Dietrich. 1978. "Einige Beobachtungen zum Sprachgebrauch von Natio im Mittelalterlichen Latein mit Ausblicken auf das Neuhochdeutsche Fremdwort 'Nation.'" Pp. 63-101 in *Aspekte der Nationenbildung im Mittelalter*, edited by Helmut Beumann and Schröder. Sigmaringen: J. Thorbecke.
- Kampinga, Herman. [1917] 1980. *De opvattingen over onze oudere vaderlandsche geschiedenis bij de hollandsche historici der XVIe en XVIIe eeuw*. Utrecht: HES Publishers.
- Kantorowicz, Ernst H. 1957. *The King's Two Bodies: A Study in Mediaeval Political Theology*. Princeton, N.J.: Princeton University Press.
- Kedourie, Elie. 1993. *Nationalism*, 4th rev. exp. Oxford: Basic Blackwell.
- Kemiläinen, Aira. 1964. *Nationalism: Problems. Concerning the Word, the Concept and Classification*. Jyväskylä: Kustantajat Publishers.
- Kist, N. C. 1848-49. *Neêrland's bededagen en biddagsbrieven*. Leiden: S. en J. Luchtmans.
- Knuttel, W. P. C. 1889-1920. *Catalogus van de pamfletten-verzameling berustende in de Koninklijke Bibliotheek*. The Hague: Algemeene Landsdrukkerij.
- Kohn, Hans. (1944) 1967. *The Idea of Nationalism*. New York: Collier.
- Kossmann, E. H. 1960. *Politieke theorie in het zeventiende-eeuwse Nederland*. Verhandelingen der Koninklijke Nederlandse Akademie Vann Wetenschappen, Afd. Letterkunde, vol. 67, new ser., no. 2. Amsterdam: Noord-Hollandsche Uitgevers Maatschappij.
- Kuiper, E. T., ed. 1924-25. *Het geuzenliedboek naar de oude drukken*. Zutphen: W. J. Thieme & Cie.
- Kunst, Christiane. 1994. *Römische Tradition und englische Politik*. Zürich: Georg Olms Verlag.

- Loon, Gerard van. 1723–31. *Beschryving der nederlandsche historiepenningen*. The Hague: Christiaan van Lom, Isaac Vailant, Pieter Gosse, Rutgert Alberts, en Pieter de Hondt.
- Marcu, E. D. 1976. *Sixteenth Century Nationalism*. New York: Abaris Books.
- Markoff, John. 1996. *The Abolition of Feudalism: Peasants, Lords, and Legislators in the French Revolution*. University Park: Pennsylvania State University Press.
- Mathiez, Albert. 1904. *Les origines des cultes révolutionnaires (1789–1792)*. Paris: G. Bellais.
- McKenna, John W. 1982. "How God Became an Englishman." Pp. 25–43 in *Tudor Rule and Revolution*, edited by Deloyd J. Guth and John W. McKenna. Cambridge: Cambridge University Press.
- Molewijk, G. C. 1988. "Een geschiedvervalsing uit 1610." Pp. 11–36 in *De oudheid van de Bataafse nu Hollandse Republiek*, edited by G. C. Molewijk. Weesp: Uitgeverij Heureka.
- Moorhead, James H. 1994. "The American Israel: Protestant Tribalism and Universal Mission." Pp. 145–66 in *Many Are Chosen: Divine Election and Western Nationalism*, edited by William R. Hutchinson and Hartmut Lehmann. Cambridge, Mass.: Harvard University Press.
- Muller, J. W. 1931. "Hooft's Baeto." *Tijdschrift voor nederlandsche taal- en letterkunde* 50:54–84, 133–72, 241–84.
- Nobbs, Douglas. 1938. *Theocracy and Toleration. A Study of the Disputes in Dutch Calvinism from 1600–1650*. Cambridge: Cambridge University Press.
- O'Brien, Conor Cruise. 1988. *God Land: Reflections on Religion and Nationalism*. Cambridge, Mass.: Harvard University Press.
- Oestreich, Gerhard. 1969. "Die Idee des religiösen Bundes und die Lehre vom Staatsvertrag." Pp. 157–78 in *Geist und Gestalt des frühmodernen Staates*, edited by Gehard Oestreich. Berlin: Duncker & Humblot.
- Ozouf, Mona. 1988. *Festivals and the French Revolution*. Cambridge, Mass.: Harvard University Press.
- Palmer, R. R. 1959–64. *The Age of the Democratic Revolution: A Political History of Europe and American, 1760–1800*. Princeton, N.J.: Princeton University Press.
- Pange, Jean Comte, de. 1949. *Le Roi très chrétien*. Paris: A. Fayard.
- Parker, Geoffrey. 1985. *The Dutch Revolt*. London: Peregrin.
- Parker, Harold Talbot. 1965. *The Cult of Antiquity and the French Revolutionaries: A Study in the Development of the Revolutionary Spirit*. New York: Octagon Books.
- Parsons, A. E. 1929. "The Trojan Legend in England." *Modern Language Review* 24 (3): 17.
- Pincus, Steven C. A. 1996. *Protestantism and Patriotism: Ideologies and the Making of English Foreign Policy, 1650–1688*. Cambridge: Cambridge University Press.
- Pocock, J. G. A. 1975. *The Machiavellian Moment: Florentine Political Thought and the Atlantic Republican Tradition*. Princeton: Princeton University Press.
- . 1987. *The Ancient Constitution and the Feudal Law: A Study of English Historical Thought in the Seventeenth Century*. Cambridge: Cambridge University Press.
- Poelhekke, Jan Joseph. 1973. *Geen blijder maer in tachtigh jaer. Verspreide studiën over de crisisperiode 1648–1651*. Zutphen: De Walburg Pers.
- Ranum, Orest, ed. 1975. *National Consciousness, History and Political Culture in Early Modern Europe*. Baltimore: Johns Hopkins University Press.
- Reynolds, Susan. 1997. *Kingdoms and Communities in Western Europe 900–1300*, 2d ed. Oxford: Oxford University Press.
- Roorda, D. J. 1961. *Partij en factie. De oproeren van 1672 in de steden van Holland en Zeeland. Een krachtmeting tussen partijen en facties*. Groningen: J. B. Wolters.
- Rowen, Herbert. 1988. *The Princes of Orange*. Cambridge: Cambridge University Press.

- Russell, Conrad. 1979. *Parliaments and English Politics, 1621–1629*. Oxford: Oxford University Press.
- . 1990. *The Causes of the English Civil War*. Oxford: Oxford University Press.
- Sahlins, Peter. 1989. *Boundaries: The Making of France and Spain in the Pyrenees*. Berkeley and Los Angeles: University of California Press.
- Salmon, J. H. M. 1991. "Seneca and Tacitus in Jacobean England." Pp. 169–88 in *The Mental World of the Jacobean Court*, edited by Linda Levy Peck. Cambridge: Cambridge University Press.
- Sande, Johan van, den. 1650. *Kort Begrijp der Nederlandtsche Historien*. Amsterdam: Ioost Hartgers.
- Schama, Simon. 1988. *The Embarrassment of Riches: An Interpretation of Dutch Culture in the Golden Age*. Berkeley and Los Angeles: University of California Press.
- . 1989. *Citizens: A Chronicle of the French Revolution*. New York: Knopf.
- Schmugge, Ludwig. 1982. "Über 'Nationale' Vorurteile Im Mittelalter." *Deutsches Archiv für Erforschung des Mittelalters* 38:439–59.
- Schöffers, I. 1981. "The Batavian Myth during the Sixteenth and Seventeenth Centuries." Pp. 85–108 in *Geschiedschrijving in Nederland*, vol. 2. Edited by P. A. M. Geurts and A. E. M. Janssen. The Hague: Nijhoff.
- Schrijver, Pieter. 1646. *Beschrijvinge Van Ovt Batavien*. Amsterdam: Broer Iansz.
- Smith, Anthony D. 1986. *The Ethnic Origins of Nations*. Oxford: Basil Blackwell.
- . 1991. *National Identity*. Reno and Las Vegas: University of Nevada Press.
- . 1994. "The Problem of National Identity: Ancient, Medieval and Modern?" *Ethnic and Racial Studies* 17:375–99.
- . 1995. "National Identities: Modern and Medieval?" Pp. 21–43 in *Concepts of National Identity in the Middle Ages*, edited by Simon Forde, Lesley Johnson, and Alan V. Murray. Leeds: University of Leeds Press.
- Smitskamp, J. W. 1947. *Calvinistisch nationaal besef vóór het midden der 17de eeuw*. The Hague: D. A. Daamen.
- Smuts, Malcolm. 1993. "Court-Centred Politics and the Uses of Roman Historians, c. 1590–1630." Pp. 21–43 in *Culture and Politics in Early Stuart England*, edited by Kevin Sharpe and Peter Lake. Stanford, Calif.: Stanford University Press.
- Soboul, Albert. 1957. "Sentiment religieux et cultes populaires pendant la révolution: saintes patriotes et martyrs de la liberté." *Annales historiques de la révolution française* 29:193–213.
- Tackett, Timothy. 1986. *Religion, Revolution and Regional Culture in Eighteenth-Century France: The Ecclesiastical Oath of 1791*. Princeton, N.J.: Princeton University Press.
- Tal, Tobias. 1898. *Oranjabloesems: Uit de gedenkbladen van neerlands Israël*. Amsterdam: Van Creveld.
- Teellinck, Willem. 1621a. *Weech-Schale Des Heylichdoms*. Middelburg: Hans van der Hellen.
- . 1621b. *Zion's Basuyne*. Middelburg: Hans van der Hellen.
- . 1624a. *Godes Handt Ter Straffe Ende Ter Velossinghe: Nu Onlangs Door Driederley Plagen in Dese Landen Gesien*. Middelburg: Hans van der Hellen.
- . 1624b. *Vvraeck-Sweet, Bepleytende Het Recht Van Gods Verbondt Door Bloedighe Oorloghe, Diere Tijd, Bleecke Pestilentie*. Middelburg: Hans van der Hellen.
- . 1650. *Den Politycken Christen. Ofte, Instructie Voor Alle Hooge en Leege Staetspersonen. Wijselijck Voorghestelt, Door Den Conincklijcken Prophete David, Inden 101 Psalm . . .*, edited by Maximillaen Teellinck. Middelburg: Anthony de Later.
- Tex, Jan den. 1973. *Oldenbarnevelt*. Cambridge: Cambridge University Press.
- Tilly, Charles. 1964. *The Vendée*. Cambridge, Mass.: Harvard University Press.

- Tilmans, Karin. 1992. *Historiography and Humanism in Holland in the Age of Erasmus: Aurelius and the Divisiekronek of 1517*. Nieuwkoop: De Graaf.
- Toit, André du. 1994. "The Construction of Afrikaner Chosenness." Pp. 115–39 in *Many Are Chosen: Divine Election and Western Nationalism*, edited by William R. Hutchinsson and Hartmut Lehmann. Cambridge, Mass.: Harvard University Press.
- Turville-Petre, Thorlac. 1996. *England the Nation: Language, Literature and National Identity, 1290–1340*. Oxford: Clarendon Press.
- Tyacke, Nicholas. 1986. *Anti-Calvinists: The Rise of English Arminianism, c. 1590–1640*. Oxford and New York: Clarendon Press and Oxford University Press.
- Udemans, Godefridum. 1640. *'T Geestelyck Roer Van 't Coopmans Schip*. . . Dordrecht: Francois Boels.
- Van Kley, Dale. 1996. *The Religious Origins of the French Revolution: From Calvin to the Civil Constitution, 1560–1791*. New Haven, Conn.: Yale University Press.
- Verwijs, E., J. Verdam, and F. A. Stoet, eds. 1882–1929. *Middelnederlandsch woordenboek*. The Hague: M. Nijhoff.
- Volk, Nation. 1972–1997. Pp. 141–431 in *Geschichtliche Grundbegriffe; Historisches Lexikon zur politisch-sozialen Sprache in Deutschland*, vol. 7. Edited by Otto Brunner, Werner Conze, and Reinhart Kosseleck. Stuttgart: E. Klett.
- Vondel, Joost van. 1937. *Volledige dichtwerken en oorspronkelijk proza*. Amsterdam: H. J. W. Becht.
- Waal, H. Van, de. 1952. *Drie eeuwen vaderlandsche geschied-uitbeelding, 1500–1800*. The Hague: M. Nijhoff.
- Walzer, Michael. 1985. *Exodus and Revolution*. New York: Basic Books.
- Watt, Tessa. 1991. *Cheap Print and Popular Piety*. Cambridge: Cambridge University Press.
- Weber, Eugen. 1976. *Peasants into Frenchmen: The Modernization of Rural France, 1870–1914*. Stanford, Calif.: Stanford University Press.
- Wootton, David, ed. 1986. *Divine Right and Democracy*. Middlesex: Penguin.
- Worden, Blair. 1982. "Classical Republicanism and the Puritan Revolution." Pp. 182–200 in *History and Imagination. Essays in Honor of H.R. Trevor-Roper*, edited by Jones Lloyd, Hugh, Valerie Pearl, and Blair Worden. New York: Holmes & Meier.
- Zagorin, Perez. 1966. *A History of Political Thought in the English Revolution*. New York: Humanities Press.
- Zernatto, Guido. 1944. "Nation: The History of a Word." *Review of Politics* 6:351–66.
- Zwicker, Stephen N. 1988. "England, Israel, and the Triumph of Roman Virtue." Pp. 37–64 in *Millenarianism and Messianism in English Literature and Thought, 1650–1800*, edited by C. Popkin. Leiden: Brill.

Book Reviews

How to Have a Theory in an Epidemic: Cultural Chronicles of AIDS. By Paula A. Treichler. Durham, N.C.: Duke University Press, 1999. Pp. xi+477. \$64.95 (cloth); \$22.95 (paper).

Lisa Jean Moore

College of Staten Island and City University of New York

Paula A. Treichler, an invaluable contributor to the fields of cultural studies, gender studies, and science, technology, and medicine studies, has coined several "sensitizing" (Herbert Blumer, *Symbolic Interactionism: Perspective and Method* [Prentice Hall, 1969]) terms such as an "epidemic of signification" to refer to the power of AIDS discourses to generate and manipulate heterogeneous signs and symbols. Primarily concerned with how knowledge and truth about AIDS are produced and represented to different audiences, she is unrelenting in her interrogation of myriad sites that produce "facts" about AIDS, such as made-for-television movies, soap operas, public service announcements, scientific meetings' proceedings, newspaper clippings, Web sites, community-based organizations' flyers, and personal observations. In part, an exploration into the dominance of science, medicine, and statistics in attempting to control, contain, or quantify the epidemic, *How to Have a Theory in an Epidemic* does not provide one alternate singular, standard, and universal explanation of AIDS/HIV. Rather Treichler's voice emerges from the cacophony of competing AIDS remediations, and without drowning them out, she guides us through understanding their historical significance and varied implications.

To a great degree, this book demonstrates how the rise of quantitative and biomedical interpretations of AIDS/HIV dwarfed, ridiculed, or eliminated qualitative and sociocultural interpretations. Ironically, it is likely that some sociologists, policy analysts, and practitioners will be uncomfortable with Treichler's initial claims and concise methodological narration. But her work beautifully makes good on the promises of theory to "develop policy" and "articulate long-term social needs even as we acknowledge the urgency of the AIDS crisis and try to satisfy its relentless demand for immediate action" (p. 1). By identifying and interpreting key historically situated texts and representations, she offers case studies. Each case study is broken down into chapters including biomedical discourse, gender, and representation; Third World representations; media constructions; Africa and cultural theory; and treatment and activism. Visual and linguistic "data" are analyzed for how they enact meanings that justify existing power structures and constitute new ones. "We cannot look 'through' language to determine what AIDS 'really' is" (p. 11). Rather, we can begin to investigate how each rendering of AIDS is the

result of a hard-fought struggle to establish the dominance of a particular system of meaning. Treichler demonstrates repeatedly how these struggles over meaning have had far-reaching implications for the misrepresentation of identities, particularly for homosexual men and non-Western people, and the erasure of groups, such as women. These misrepresentations and erasures have contributed to the ineffectual execution of ill-conceived AIDS education and prevention campaigns.

Part of what is so ingenious about her work is that while Treichler uses cultural theories to deconstruct and analyze images and texts about AIDS, she does not ignore the material, physical, or somatic actuality of AIDS. In practical terms, it is extremely difficult to constantly problematize the construction of truth of AIDS, while human bodies are in pain, wasting away, or dying. The chapter "AIDS, HIV and the Cultural Construction of Reality" is a superb contribution to the sociology of knowledge, in that it traces different constructionists' views of reality. We must see how reality and our linguistic interpretation of it are co-constitutive. This review of theory is accompanied by an excellent example that exposes the "fabrication and cultural constructedness" (p. 168) of a virus.

Another significant contribution of this work is the way in which cultural theories are used to make visible the contested nature of scientific knowledge. Within the context of AZT (zidovudine), an AIDS treatment drug, Treichler explores how interactions between activists, performers, and scientists create a space to witness the ruptures and fractures in scientific knowledge. Throughout the book, we learn how knowledge production is a multifaceted and multitiered social process that happens within existing social relations. The construction of what qualifies as knowledge is "nonarbitrary"; it is a dynamic process laden with unspoken agendas and unanticipated resistances.

Although occasionally the writing is dense, it is clearly due to the challenge of maintaining Treichler's enterprise: using theory to understand the "structural and cultural characteristics that promote the generation of meanings" (p. 316). The use of visual reproductions, elaborate endnotes, and a thorough bibliography greatly enhances readers' ability to utilize this work. Furthermore, readers are likely to appreciate Treichler's keen sense of humor peppered throughout in the form of parenthetical comments and asides. This book is a welcome addition to any syllabus related to medicine; science; the sociology of knowledge; the media; social movements; and gender, race, class, and ethnicity. While each chapter is coherent and could stand alone, readers best experience the magnitude and power through reading the entire contents. Indeed, sociologists, anthropologists, historians, and literary and media critics, as well as epidemiologists and clinicians are fortunate to have such a blessing as Treichler's extensive research and interpretation of AIDS/HIV.

Lessons from the Intersexed. By Suzanne J. Kessler. New Brunswick, N.J.: Rutgers University Press, 1998. Pp. x+193. \$48.00 (cloth); \$18.00 (paper).

Bernice L. Hausman

Virginia Polytechnic Institute and State University

The cover of Suzanne Kessler's *Lessons from the Intersexed* displays Norman Rockwell's *Visit to a Family Doctor*, an illustration that conveys the book's main themes: the physician's authority to give expert advice to naive parents, the parents' concern about their child, and, most significantly, the baby's initial sexlessness. Inside, the book packs a wallop of an argument: current treatment protocols for intersexed infants are unethical because they (1) emphasize the "normal" appearance of genitals rather than their urinary and sexual functioning, (2) are based on phallogocentric standards that insist that clitorides really should not be seen and penises must be of a certain length, (3) encourage parents to make hasty, isolated decisions while the infants are still very young, (4) often create more anatomical and physiological problems than were present at the infant's birth, (5) psychologically damage intersexed persons through repeated hospitalizations during childhood and the aura of secrecy that often attends the surgical procedures, and (6) promote the connections between genitals and gender that our society depends upon to make sense of persons, identities, and social and sexual roles.

A balanced, highly readable, and fascinating book, *Lessons from the Intersexed* enters into current feminist debates about gender as a study of a specific phenomenon of sex that confounds the commonsense understanding that humans come in only two types, and concludes that we must sever gender's relation to the biological body and allow it to become an always in-process aspect of the "social-interactional body" (p. 132). Kessler's discussion is based on interviews with adult intersexuals, parents of intersexuals, and physicians who treat intersexed infants; she also engages debates about gender in feminist, gay, and transgender scholarship. The greatest value of *Learning from the Intersexed* is that in it Kessler demonstrates how ideologies of gender work in the real world to configure sex.

Current intersex treatment protocols—which mandate invasive procedures to "correct" ambiguous genitalia early in the infant's life—make sense only in the context of a strict belief in the necessity of sexual dimorphism for culture's smooth operation, both for the intersexed individual and for everyone else. The belief, however, is masked by the assumption that sexual dimorphism is an obvious social requirement. Thus the medical profession, Kessler argues, ignores evidence of poor surgical outcomes, subjugating such data to their "commitment to the concept of medical advancement and that of dimorphic genitals" (p. 74). As a result, physicians seem to be offering the most viable solution to their intersexed patients; alternatives simply are not part of the conceptual landscape. Yet

the experiences of both parents of intersex infants and adult intersexuals demonstrate that while "physicians pride themselves on shielding parents from ticklish gender issues, . . . the confrontation is only delayed [for the intersexed subject]" (p. 100), as he or she tries to make sense, as an adult, of confusing and highly charged childhood experiences around the meaning of his or her body.

One problem with *Lessons from the Intersexed*, however, is Kessler's use of the word *gender* to designate both biological sex attributes and cultural articulations of gender (simplistically, masculinity and femininity). While I agree that "biological sex" is a concept that has a contingent relation to what we might think of as "nature," I find it problematic not to distinguish between the socially constructed concepts of gender and sex. Using the word *gender* to underscore the constructed nature of sex only serves to confuse certain arguments more, because the reader does not always know the specific register to which Kessler refers. And the argument breaks down, because at the end of the book, she writes, "Once we dispense with 'sex' and acknowledge gender as located in the social-interactional body, it will be easier to treat it as a work-in-progress" (p. 132). Using the word *sex* with scare quotes underscores Kessler's discomfort with the category, but she has to use it because she is making an argument in which a distinction between sex (biological genital dimorphism) and gender (performative body in culture) is crucial—even though it is precisely this distinction she has denied through her inclusive use of the term *gender*. Throughout the book a more profitable approach would be to use both sex and gender as terms to reference distinct constructed concepts.

But really, that is just a theoretical quibble. As Kessler aptly shows, intersexed infants are subject to the gender system at its most ruthless—their bodies are literally carved up in accordance with the culture's gender mandate. Rockwell's painting on the cover emphasizes gender's claim on all of us, reminding us that it is a social, and not bodily, imperative by portraying the familial and commonplace contexts in which gender is a salient and seemingly necessary aspect of social relations: the mother holds the baby, the physician's authority rests partially on his being an older man. This baby must, for all intents and purposes, become a boy or a girl.

Epistemic Cultures: How the Sciences Make Knowledge. By Karin Knorr-Cetina. Cambridge Mass.: Harvard University Press, 1999. Pp. vii+329. \$22.95 (paper).

Trevor Pinch
Cornell University

Karin Knorr Cetina is a pioneer of the genre of sociology of science known as "lab studies." Her 1981 book, *The Manufacture of Knowledge*,

was one of the first ethnographies of a scientific laboratory. Like most work at the time, she was concerned with showing how particular knowledge claims and entities in science were constituted from the social, discursive, and material practices to be found in laboratories. The claim that scientific knowledge is "manufactured" or "constructed" stimulated much debate—a debate that still lingers on today in the form of the so-called "science wars." Her new book, which examines molecular biology (MB) and high energy physics (HEP) is less likely to provoke scandal. She is not primarily concerned with understanding how scientific knowledge is socially constructed, rather, she compares how these two scientific fields are organized and their different strategies for acquiring knowledge. This book is, in short, the first ever comparative laboratory study.

That HEP and MB are organized in very different ways is scarcely news. Because of the cost and size of the accelerators used in HEP there are just a few facilities world wide. Investigating particles created at gigantically large energies, which exist for only minuscule moments of time, makes the endeavor extraordinarily complex. It involves teams of physicists from many different labs, who plan for years in advance how they will come together for a particular experiment. The most obvious manifestation of the peculiarity of the field is the numbers of authors listed on scientific papers—this can often run into several hundreds. MB is dispersed between far more smaller labs and correspondingly smaller teams of scientists who often work in competition with each other, each under the leadership of one laboratory head.

One of the most remarkable observations in the book is that physicists for long periods of time engage in a form of self-reflection and self-analysis of the way that they study objects. Natural and quasi-natural objects are admitted to experiments only rarely. Physicists deal with highly interpreted entities—they are immersed in a sea of signs. Their interpretative practices are guided by theoretical knowledge, models, simulations, and statistical procedures. Above all they must know what they do not yet know. No stone is left unturned; every nuance and quirk of their detectors are minutely examined; background effects, noise, and distortions must be accounted for and "beaten down." They study what Knorr-Cetina aptly calls "liminal" phenomena. By focusing on "negative knowledge," on the multifarious ways in which things can go wrong, HEP manages to produce positive knowledge. Knorr Cetina intriguingly compares this approach with what is known as "apophantic theology" in Christianity—studying God by what He is *not* rather than by what He is (p. 64).

In contrast, the molecular biologists know God by his works. MB works by maximizing interaction with natural objects and puts a premium on the acquisition of experiential knowledge. If HEP is characterized by self-referential, negative epistemics, built around sign systems, MB seeks positive knowledge to be gained from the manipulation of material objects in an experimental regime that tries to look through the signs for clues as to the material reality behind them. The different strategies are revealed when problems are encountered. Molecular biologists

respond to a problem, such as a malfunctioning reaction, not as the physicists do, by doing everything possible to try and understand the problem, but rather by trying different variations of their procedures in the hope that this will lead to new evidence. Blind variation pays off as a strategy when you have a vast array of small objects, many intervening technologies, and an emphasis placed on an experiential empirical reality.

Sometimes the results of the comparison between HEP and MB are surprising. Knorr-Cetina finds that physicists describe their detectors in terms usually reserved for physiological organisms—they talk about their detectors “aging” and “playing up.” The molecular biologists, on the other hand, describe their actual living organisms with a machine-like vocabulary: a mouse is not an animal but an “experimental device.”

HEP involves a “post-traditional communitarian structure,” which uses “management by content.” This is a form of distributed cognition that keeps the many different sorts of practitioners close to the source of their expertise and the objects they work with while at the same time providing a social structure that keeps the complex as a whole operating. “Leaders,” who are more akin to spokespeople, are not at the top of the experiment, but rather centrally located within a conversation conducted within a collaboration. MB is much more traditionally organized, with each lab having a hierarchy dominated by one leader who *de facto* is the head and spokesperson for that lab. The different organizational form is used to explain why collaboration is more typical of HEP and competition of MB.

Knorr-Cetina wants to tie this study to the work of social theorists who argue that Western societies are ruled by “knowledge societies,” “risk societies,” “information societies,” and the like. It is certainly the case that social theorists have much to learn from the sociology of science, but the problem is that it is not obvious why the particular organizational and epistemic forms found in these two sciences should be typical of the organization of knowledge and expertise outside the sciences. One can imagine an equally ambitious project for which Knorr-Cetina has laid the groundwork: namely a comparative project in epistemics and organizations for the sciences as a whole. Even if one is not that ambitious, one could ask further questions: Is MB more typical or HEP? Is sociology closer to MB than HEP? What other forms of epistemics and organizations are possible? Even more interesting is to relate the old question of how knowledge claims are constituted to the different epistemics and organizational forms she has documented. This is a topic she touches upon in the conclusion (which consists of a dialogue between a fictional reader and the author), where she suggests that controversy over knowledge claims is less likely in HEP because of the way that time is reordered, such that well before an experiment is carried out the collaborations must be in place. This makes it less likely for parts of the collaboration to break out in disagreement later on. This groundbreaking book will not only be of interest to sociologists of late modernity, but also should rekindle interest in the comparative sociology of science.

A Structural Theory of Social Influence. By Noah E. Friedkin. New York: Cambridge University Press, 1998. Pp. xix+231. \$59.95.

Stanley Wasserman
University of Illinois

It is a pleasure to review this monograph for the *American Journal of Sociology*, not only because of its high quality, but also because reviews of similar texts are getting hard to find. Sociology journals appear not to be reviewing most of the newly published methodology books, as well as those treatises whose main foundations are quantitative. *A Structural Theory of Social Influence* falls into the latter category and presents Noah Friedkin's research over the past decade on social influence models for social networks in a substantive framework. The basic premise of this research is that large-scale social systems have inherent social differentiation that gives rise to different standards, opinions, and values. The actors in such systems are influenced by the diverse characteristics and beliefs of the actors with whom they communicate, are friendly with, work with, and so forth. But network models are ill-prepared to handle such complexity.

Friedkin proposes a *structural social psychology* that can operate in a complexly differentiated group but, unlike standard structural analyses, can synthesize all the information contained in a *complex* social network. Such information not only consists of a set (or sets) of actors, relational measurements on these actors (which Linton Freeman views as constituting the "algebraic structure" of the network), but also attributes or characteristics of the actors, such as age, gender, attitudes, and so forth (often called "standard" social/behavioral scientific measurements). The focus of Friedkin's monograph is how one (or sometimes more than one) attribute(s) can be "predicted" by relational information, as well as by other attributes. Structural analyses of social networks often produce partitions of actors into social positions that can be studied more easily than the full collection of individual actors. Unfortunately, this type of structural analysis, quite standard in the study of social networks, lacks formal theories that help predict these social positions of actors, based on relational information. It is easy to find blockmodels (see the classic work by Harrison White and colleagues) or cluster analyses of actors, both of which produce actor partitions, but it is difficult to statistically determine why such actor partitions arise (but see Arabie, *Social Networks* [6 (1984): 373-403]) or how they are associated with measured, evolving opinions or behaviors. What has long been missing in social network analysis are theories and techniques to relate structural analyses to these measured attributes.

Friedkin works toward a formal model of an entire social system. He illustrates his approach using data from faculty members, from a variety of departments, at two universities. The data analysis, as well as the substantive motivations for the analyzed measurements, constitutes nearly

two-thirds of the monograph, making the text a wonderful adventure in modeling, both substantive and quantitative, and analysis.

Friedkin's model for *social influence*, also called the network effects or social process model, provides a basis through which one network actor alters the behavior or attitudes of another. Substantive foundations of social influence include the work of Leon Festinger, Georg Simmel, and Robert Merton, but especially Émile Durkheim, to whom Friedkin pays much tribute. Specifically, Durkheim's *Division of Labor in Society* (1933) supplies Friedkin with arguments stating that actors in corporate organizations seek other actors with similar attitudes and opinions. Such cohesion leads to better running organizations; furthermore, the cohesive subgroup memberships should be predictable from relational information. Friedkin states: "While social structural analysis allows a description of complexly differentiated social structures, it has failed to develop a theory that elucidates the consequences of such structures for the production of interpersonal influences and agreements" (p. 23). Models such as these nicely integrate many of the dominant aspects of social network analysis, especially the statistical and the algebraic. Indeed, it is remarkable how little effort has gone into work on a merger of these two dominant network methodologies of the 1970s: the statistical models of Davis, Leinhardt, and Holland, which for the first time, allowed a researcher to evaluate complex hypotheses about social structure using sophisticated probability distributions, and the relational algebras and blockmodels of White and colleagues.

The social influence model has two basic equations. One assumes that the opinions/attitudes/attributes of actors are linear functions of a collection of "exogenous" variables, while the other (which brings in relational information) assumes that the opinions measured at time t depend on the sum of opinions at time $t - 1$ and the initial opinions (time 0), weighted appropriately. The opinions at the earlier time point are also weighted by "endogenous interpersonal influences" for each actor (through which enters the relational measurements). Special cases of this basic model include statistical models arising in spatial econometrics that (unfortunately) include error terms (often assumed to be normally distributed). The exogenous variables, and calculations of the endogenous network variables, are described at length in chapters 4–7, woven with illustrations taken from the departmental data sets. Interested readers should check out more recent social influence models by Robbins (*Psychometrica*, in press) that should make earlier approaches obsolete.

The book is not an "easy read," but the mathematically adventurous will be rewarded with a fresh look at complicated networks, and with avant garde tools that will be very useful in practice. Those not so adventurous will find easier reading in the more substantive chapters, which do stand alone from the mathematical presentations.

On Line and On Paper: Visual Representations, Visual Culture, and Computer Graphics in Design Engineering. By Kathryn Henderson. Cambridge, Mass.: MIT Press, 1999. Pp. 243. \$30.00.

John Law
Lancaster University

How do people *think*? How do they conceptualize problems? How do they design objects?

Kathryn Henderson's book is a study of design. In particular, it is a study of the way in which engineers make use of visual materials in the process of engineering design. Her thesis is that design does not simply take visual form before being converted into a microchip or a turbine, but that the whole process is visual as well as verbal and cognitive, through and through.

So far so good. But how to think about the role of visualization in design? Here she draws on and extends findings of the discipline of science, technology, and society. The argument is that visualizations do not appear in mid-air. They need to take some *material* form. This sounds like a trivially obvious point, but it has less than obvious implications. This is because the forms of visualization, the media into which they are inserted, and the cultures or visual languages that evolve around and through those media are all highly specific. They are "forms of life," professional and technical practices embedded in the understandings, the tacit knowledge, and the skills of engineers and designers. They enable communication between disparate groups. And sometimes, inevitably, they are also used in power play—between or within organizations and professional groups.

Henderson's book is a careful exploration of the "visual languages" of technical practice. As a part of this, she explores the evolution of engineering standards in the United States and its links with the development of professionalization, industrial self-regulation, and government intervention. She also develops her analysis of contemporary design practices through a series of interviews. But it is ethnography that lies at the heart of the book. The author describes two participant-observation studies in industrial design settings. Using her skills as a professional trained both in art and sociology, she spent a year working as a technical writer in a firm producing turbine engines. Then she moved on to participate in the work of a company that makes surgical instruments. And the argument of the book grows predominantly from these two ethnographies.

These two case studies allow Henderson to approach the core topic of the book, which is the advent of computer-aided design (CAD) in engineering. For with the growth of information technology, there has been a revolution, or at least half a revolution, in engineering practice over the last 20 years. In the past, engineers worked through drawings—back-of-the-envelope sketches, plans, profiles, and sections and lists, schematics (like wiring diagrams), and illustrative cutaway diagrams. These were

the visual languages of design practice as products moved from the drawing board to the detail design to the mock-up to the working model to the final production version (though this "yellow brick road," the linear model of engineering design, as Henderson shows, was never anything other than a chimera). Now, however, there are the new CAD tools that promise new forms of visualization, new economies, new forms of standardization, and novel methods of communication. But these changes, as Henderson carefully shows, though they bring advantages at the same time, bring with them costs that come in several forms. For instance, many are unfamiliar with the new visual syntaxes of CAD and cannot use them or interpret them. Those syntaxes themselves are in some cases very different from those of paper-based drafting—they produce quite different visualizations, and sometimes these simply look like errors in the language of the paper-based drawing board. Again, though CAD is powerful in many ways, in others it is inflexible—empirical methods of drawing approximation may, for instance, be beyond its reach. And it may further divide the world of drawing from the shop floor world of practice.

Henderson's book is beautifully researched, located as it is in the rich tradition of symbolic interactionist ethnography. However, its insights draw on a wide range of insights from science, technology and society, art history, and engineering history. It is an important book in many respects, and not least because it shows how it is possible to avoid the extremes that attend many of the current commentaries on information technology. Slipping between hype on the one hand and Luddite gloom on the other, it shows how it is possible to conduct theoretically well informed and empirically incisive work on contemporary technologies in a way that is both sympathetic and critical. And, as I noted at the beginning, it tells us much about how people think in a material context. I recommend it highly.

Impersonal Influence: How Perceptions of Mass Collectives Affect Political Attitudes. By Diana C. Mutz. New York: Cambridge University Press, 1998. Pp. xx+334. \$54.95 (cloth); \$19.95 (paper).

Lynn M. Sanders
University of Virginia

Impersonal influence refers to a seemingly incontestable social fact: increasingly, people gather information about the world through abstract representations of collectivities. At the core of this book is a set of results from psychological experiments and surveys, testing in various ways the proposition that political thinking is influenced by people's perceptions of collectivities that are represented and imagined, not experienced. Such impersonal entities include nations, states, some groups (e.g., "doctors" or "working women") or concepts ("the economy" or "public opinion"). Since

knowledge of such abstractions is necessarily mediated, as personal exchange grows less frequent and the mass media more prominent, impersonal influence stands to be an increasingly potent force.

Mutz is interested in how collective representations compare to direct experience when people are assessing political problems or making political choices. Because there might be special opportunities for coercion or misrepresentation when people make political generalizations about collectivities to which they do not actually belong, the animating normative question of this book is about democratic accountability. Though Mutz began her research thinking she was studying a "pernicious phenomenon" (p. xv), as she pursued it, she changed her mind. She concludes that impersonal forms of influence in fact enhance democracy, by exposing people to more experiences and points of view than they could possibly encounter on their own.

Mutz's evidence justifies this conclusion. People appear to use representations of collectivities in judicious and often sophisticated ways. For example, when presented with representations of public opinion, people do not act like sheep and change their views of candidates or issues to match the views other supposedly hold. Instead, Mutz demonstrates that citizens use such representations rationally to reconsider but not necessarily to abandon their own views. Abstract representations of group experience seem also to prompt sophisticated considerations by citizens. Mutz marshals evidence that citizens evaluate presidential candidates not on the basis of their own parochial interests but instead consider how various groups, including ones they do not belong to, are faring economically relative to each other.

Thus abstract representations of impersonal experience have the potential to move people outside of their immediate life-space, to make broader-based evaluations of justice and fairness. Indeed, one of the most satisfying elements of the book is its challenge to the assumption that face-to-face interactions are necessarily democratically preferable to impersonal forms of communication. Not only might the "internalized conversation" individual citizens engage in when confronted with representations of others' views and experiences adequately substitute for the face-to-face exchange that is wistfully supposed to have taken place at some unspecified point in the past. More strongly, Mutz suggests that individuals' engagement with impersonal representations is "likely to be less parochial than most people's interpersonal contacts" (p. 290).

This said, it is not clear that Mutz has dismissed, or even really addressed, the concerns of mass society theorists from Gabriel Tarde to the Frankfurt school to contemporary figures like Jean Elshtain and Robert Putnam. Their core worry concerns the decline of opportunities for face-to-face public exchange. Addressing it would seem to demand a direct comparison of how political information, both personal and impersonal, is considered by individuals in isolation, compared to face-to-face, in order to assess whether the decline of personal associations implies a decrease in the quality of political thinking.

Though Mutz's normative conclusion directly contrasts face-to-face to impersonal influence, they are not, strictly speaking, directly tested against each other in the analyses on which it is based. Mutz draws her evidence from experiments and surveys that situate individuals apart from each other. Therefore, she shows that citizens in isolation think about collective as well as personal experience in arguably quite democratic ways—but not that they are more democratic or less parochial than they would be if they considered the same questions and faced the same study conditions, only face-to-face. People may consider *both* abstract, collective, impersonal experience *and* immediate, personal experience *either* in face-to-face settings *or* in isolation from each other. Considering collective experience in isolated settings may be superior to ignoring the world beyond the personal, but thinking about collective experience in face-to-face circumstances might be better still.

To focus on Mutz's failure to test one implication of mass society theory would, however, be to miss her broader intention in identifying impersonal influence, which is "not a unitary theory so much as a collection of closely related phenomena united under this umbrella term" (p. 22). Mutz's theoretical invocations are intended to point up family resemblance, to show the relationship, for instance, between such seemingly remote entities as 19th-century French social theory and the secure political science finding that collective considerations of economic well-being outweigh personal ones when Americans vote. Indeed, this book succeeds in offering a new way of thinking about seemingly remote social science findings, both established and new. It does so in a way that is both more ambitious and more profound than most books about political psychology. For its many interesting findings and the theory that organizes them, this is a book that no student of public opinion or political communication should miss.

Identity without Selfhood: Simone de Beauvoir and Bisexuality. By Mariam Fraser. New York: Cambridge University Press, 1999. Pp. 225. \$64.95 (cloth); \$23.95 (paper).

Janet Wirth-Cauchon
Drake University

Mariam Fraser's *Identity without Selfhood* examines the narrative techniques that produce the figure of Simone de Beauvoir as an intelligible self. Fraser uses Beauvoir as a "cipher" or "vehicle" through which to explore, in particular, sexual narratives as constitutive of coherent selfhood. Drawing on Michel Foucault's concept of "techniques of the self," Fraser's aim is twofold: First, Fraser examines the discursive techniques that biographers, journalists, and academicians use to produce the "truth" of Beauvoir as an individual self, a truth that revolves around her relationship with Jean-Paul Sartre and existentialism. Second, Fraser traces

the ways bisexuality is minimized or displaced from significance in these biographical narratives of Beauvoir, effectively precluding the possibility that bisexuality was integral to Beauvoir's selfhood. In these narrative renderings of Beauvoir, her relationships with women are rarely if ever depicted as important in and of themselves, nor as arising out of Beauvoir's own independent desires for other women, but rather are shown to be inextricably tied to her relationship with Sartre, or to experimentation linked to existentialism and a bohemian lifestyle. In the rare instances when her relationships with women are depicted as arising from Beauvoir's active choice, however, they are understood to be evidence not of her bisexuality, but of a suppressed lesbian identity. The case of Simone de Beauvoir, then, presents an example of the "crisis of meaning" posed by bisexuality to essentialist conceptions of sexual identity, whether heterosexual or lesbian. (Amanda Udis-Kessler, "Present Tense: Biphobia as a Crisis of Meaning" in *Bi Any Other Name: Bisexual People Speak Out*, edited by Loraine Hutchins and Lani Kaahumanu [Alyson Books, 1991]).

Fraser's book makes a major contribution to recent scholarship in feminist, poststructural, and queer theories of subjectivity, the body, and identity by tracing this bisexual "crisis of meaning" in a specific case to expose the limits of current assumptions regarding sexual identity. The strength of Fraser's analysis is its theoretical erudition and richness, making it a must-read for scholars interested not only in theories of subjectivity, but also in narrative analysis, which is brilliantly brought together with theoretical analysis in this text. Fraser has traced, with painstaking precision and care, the discursive threads by which a coherent portrait of de Beauvoir is woven, whether as heterosexual or as lesbian. Chapters 3 and 4 use Foucault's concept of the "author-function" and Paul Ricoeur's work on narrative employment to examine biographical narratives and the specific devices through which the "truths" of Beauvoir's self are constructed and a bisexual self is precluded. Chapter 5 shifts the focus to journalistic and photographic portraits of Beauvoir, examining the ways Beauvoir is "framed" as heterosexual, with bisexuality displaced from visible representation.

If the analysis were to end here, it would have provided a detailed case study of heterosexism in biographical representation. Yet Fraser goes on to consider the ways Beauvoir's life and work are also constructed in relation to a coherent lesbian identity and to lesbian feminist politics. Here, Beauvoir's relationships with women are taken as signs of her underlying lesbian identity, which Beauvoir was unable to express, given the historical time period, or which is viewed as "inauthentic" because her lesbianism was expressed in same-sex relations that were self-serving and exploitative. In challenging these readings of Beauvoir, Fraser shows how these accounts make an explicit connection between pleasure, responsibility, and loyalty to other women as features of an "authentic" lesbian political identity. Once again, but for different reasons, bisexuality is erased.

With Beauvoir used only as a “cipher” for an analysis of Foucaultian subjectivation, readers looking for an in-depth account of Beauvoir’s biography will not find it here. Fraser provides none of her own background on Beauvoir’s life and allows such background to speak, in fragments and in footnotes, only through the texts she is examining. Beauvoir herself—her life and her work—disappears behind the narrative and discursive figure of intelligibility, which is the real aim of Fraser’s analysis and critique. This is appropriate to Fraser’s aims, for to merely offer yet another authoritative account of Beauvoir would be to reproduce the very discursive “techniques of selfhood,” which Fraser is critiquing.

It is the “work” that Fraser makes the case of Beauvoir’s bisexuality do, for contemporary debates in theories of identity and embodiment, that will be of most interest to readers. Bisexuality presents a paradox: it is “‘visible’ but does not constitute a sexual-narrative-identity which could be ascribed to a self” (p. 126). Unlike other sexual identities, it is not necessarily anchored in an embodied, material self. In an early chapter on feminist and queer theories of embodiment, and in the final chapters of her book, which draw on Deleuze and Guattari’s concept of desire, Fraser offers an alternative account of identity. Bisexuality helps “undo” the logic of identity, illustrating a more mobile, Deleuzian desire that is not tied to a fixed self.

Fighting to Become Americans: Jews, Gender, and the Anxiety of Assimilation. By Riv-Ellen Prell. Boston: Beacon Press, 1999. Pp. vii+319. \$28.50.

Deborah A. Gerson
San Francisco State University

The overbearing Jewish mother, the spoiled Jewish princess, and the ineffectual and spoiled Jewish man are fading staples of post–World War II American culture. But the imprints of these visions remain in the psyches of contemporary Jewish men and women observed by Riv-Ellen Prell in a series of dialogue workshops. Using data from the English pages of Yiddish-language newspapers, from literature, biography, and diary, Riv-Ellen Prell attempts to decode how images, constructions, and relations of gender have changed for Jews from the great wave of immigration until the 1990s. Prell explores a triangular set of interactions and intersections involving gender, Jews, and the process of Americanization. She asks why and how do Jewish men and women not get along? How does gender complicate and implicate itself in the process of group assimilation? And how does the experience of class mobility in an increasingly affluent and consumerist society shape the relations of Jewish men and women?

Prell introduces us to the ghetto girls of the Lower East Side—young women who dressed fashionably, wore makeup, and seemed to flaunt

their independence and sexuality—a “frantic, panicked rendering of young Jewish womanhood . . . created by Jews who felt threatened by non-Jews, men threatened by women, native born . . . frightened of a nation of immigrants” (p. 43). The cultural representation of the ghetto girl, like related images of other subordinated racial and economic groups, divided the very group it targeted. Jewish men and women were embarrassed by, rejecting of, and distanced from these women.

Jewish parents wanted their children to marry within the people, and mostly they did. But labor force participation by both men and women created a set of conflicting demands. On the one hand, young women wanted a man to relieve them of the pressures of wage labor, which, despite the rewards of dollars and autonomy, was tiring and often oppressive. On the other hand, women’s participation in the labor force raised the bar of acceptance of a mate: why stop working and indulging oneself if a man could not provide? Working-class men and women faced each other with fear and suspicion; wanting to rise in class and status through marriage, women might refuse the proposals of working-class men, while the men feared that women required too much wage labor from them and provided too little support in return.

Jewish gender trouble is money trouble, class trouble, and status trouble. If immigrant troubles were those of the poor and working class, with post-World War II prosperity came a different variety of gender trouble. The expanding American economy made possible Jewish class mobility, transforming the organization, structure, and purpose of Jewish family life. For Jewish male writers, the young, beautiful, and consumerist daughters, unwitting ball busters who will trap creative men, personify the “success and excess” (p. 223) of the postwar Jewish suburbs of New York. Hard work and sexual monogamy made one a good husband and father; they also diminished men’s control over time, self, and resources. Marjorie Morningstar (formerly Mongenstern) meets Noel Airman (a wonderful play on the Yiddish *luftmentsh*, literally airman and figuratively space head). The stereotypes of spoiled young women, overbearing mothers, and ineffectual or inadequately masculine men could gain social currency not because they were cogent representations of Jews, but because they crystallized the social anxieties of white middle-class American men. If conspicuous consumption marked success, it also signified the demands of the market: gendered as mother, daughter, or wife—the extravagances of consumption were named Jewish mother and Jewish princess. Prell sums up: “in the 1970s, 1980s and 1990s . . . to love and marry a Jew became equivalent of being a slave of middle-class life” (p. 208). But in a curious retreat from her gender analysis, she omits the obvious question: For whom?

Following the emergence of second-wave feminism, Jewish women created new, subversive representations of their own lives, foregrounding not their Jewishness but themselves as everywoman: affluent, self-absorbed, insecure, ironic, and as the middle initial of JAP points out, American. In fiction, on TV, in movies, and through art, Jewish women

have repackaged and co-opted the very stereotypes that diminished them. Surprisingly, Prell diminishes the implications of her own data and concludes that "the astonishing persistence of Jewish gender stereotypes serves as a reminder that winning the fight to become Americans exacts a devastating price" (p. 245).

Given that Prell provides us with multiple examples of women inverting, subverting, and resisting gender stereotypes, and if TV counts and Seinfeld is our new everyman, then I do wonder, how devastating the price? On every social indicator, Jews as a group are doing very well. While Jews (except for the Orthodox) seem to have rejected the injunction to "be fruitful and multiply," we do live long and prosper. Gender suspicion and hostility can accompany class oppression as well as class mobility, witness hip-hop as well as JAPs, with more dire consequences for group survival and well-being. Riv-Ellen Prell shows us the gender wounds sustained in the "fight to become Americans" but underestimates just how well these wounds can heal in the process of becoming, in Karen Brodtkin's terminology, American white folks (Karen Brodtkin, *How Jews Became White Folks* [Rutgers University Press, 1998]).

Handbook of the Sociology of Gender. By Janet Saltzman Chafetz. New York: Kluwer Academic/Plenum Publishers, 1999. Pp. xiii+630. \$130.00.

Myra Marx Ferree
University of Connecticut

This book is a mammoth undertaking, both for the editor and for any potential reader. As a handbook, it presumably is intended not to be read from cover to cover but rather consulted on specific topics as the need arises. It meets this reference purpose excellently, offering 27 chapters on the various aspects and dimensions of gender in society, each of which is written by a sociologist of considerable reputation in the subarea, and offers a splendidly long and comprehensive bibliography.

There are four subareas in this volume. The first covers such core and cross-cutting issues as feminist epistemologies and gender theory; the second defines itself as focused on such macrolevel issues as societal development, migration, organization, and culture; the third tackles microlevel structures and processes such as socialization and interaction; and the last and largest represents an assortment of mesolevel institutions such as sports, politics, science, healthcare, and education. This spread from micro- to macrolevel analysis represents the full scope of gender analysis in sociology today. As such, it is a valuable resource and corrective for the standard introductory sociology-type understanding of gender, which concentrates largely at the microlevel and stresses socialization and social roles, just two of the 27 topics under consideration here. (See detailed analysis of textbooks by Ferree and Hall "Rethinking Stratification from a Feminist Perspective: Gender, Race and Class in Mainstream Text-

books," *American Sociological Review*, 61 (1996): 1–22). It is to be hoped that having such a compendium of current research and theory so accessible will have an influence on how issues of sex and gender are approached in the classrooms of the future.

It is a resource that every library should have, not only for teachers, but also for their students to consult. The essays invariably orient the reader to the key issues and controversies in a subarea, and the bibliographies are extensive without attempting to cover every reference that might exist. For students, each essay will function as a useful tool to discover what one knowledgeable researcher in the area would consider important, almost like having an extra advisor to guide them through the field. It is inevitable that such choices in a large and rapidly growing field are idiosyncratic to some degree.

As vast and as valuable as this collection is, it is hard not to quibble with some of the editor's specific choices about what to emphasize, particularly in the extensive (13 chapters) section on social institutions. It is clearly impossible to cover all social institutions in all regards equally well, but there is an odd mix of narrow and wide topics that makes some of the bibliographies focus in specific directions that may be more or less useful to the generalist who turns to this volume for an overview. For example, the only essays with a specific focus on the intersection of race and gender are the chapters by Barbara Reskin and Irene Padavic on U.S. workplaces and the one on gender and international migration by Pierrette Hondagneu-Sotelo and Cynthia Cranford. In most chapters, such as the one on gender and science, the statistics offered take women as a homogeneous whole and thus obscure the specific situation of women of color in these institutions. The intersection of systems of class, race, and gender as such scarcely appears at all, though it would have seemed more naturally a part of a section on macrostructures and processes than the chapter on social movements that is found there. The study of social movements would seem to belong among other social institutions instead, particularly as a counterpoint to institutional politics. Conversely, the essay on gender and politics assumes a narrow, gendered definition of politics as limited to the formal institutions of voting and office holding rather than encompassing social movements, community organizing, and other forms of grassroots political activity where women are more likely to be seen.

Yet, despite these and other small reservations about the organization and the focus of the specific chapters, I find the volume as a whole to offer a broad range of diverse sources and perspectives that are conveniently brought together in a single compendium. For a handbook that necessarily presumes a certain level of disciplinary knowledge, and hence some degree of jargon, it is relatively accessible and clearly written, which is a testimony not only to the authors but to the editor. This book is a substantial resource (albeit at a substantial price) that libraries serving sociologists should be sure to include among their holdings.

Equality by Design: The Grand Experiment in Destratification in Socialist Hungary. By Szojna Szelényi in collaboration with Karen Aschaffenburg, Mariko Lin Chang, and Winifred Poster. Stanford, Calif.: Stanford University Press, 1999. Pp. 267. \$45.00.

Petr Mateju

Academy of Sciences of the Czech Republic

There are two types of new book in the field of sociology that one finds useful to read. First of all, there are intellectually provocative books, challenging old theories and revealing pathways leading to new ones. In this type of book, usually written by "intellectual leaders" of the discipline, empirical evidence, if present at all, is used to illustrate new ideas rather than to prove them. To bring this a step further, this kind of intellectual challenge must be taken up by "workers" of the discipline, who follow the pathways and pave them with stones shaped by rigorous analyses of empirical material from solid social research. These "workers" of the discipline are just as capable of producing a book one really enjoys reading provided they do their job both with the same passion and care as the intellectual leaders. The book being reviewed belongs to this category.

In particular, this is due to the author's decision to challenge, on empirical grounds, a number of conventional views on the grand experiment in destratification in former socialist countries. She also debunks some myths regarding the formation of elites during the postsocialist transformation, although this latter area is, unfortunately, only partially covered. After a short review of the main approaches to egalitarian reforms, the real achievements of the egalitarian project that took place in socialist Hungary are evaluated in four chapters: "Schooling for Socialism"; "The Class Structure of Classless Hungary"; "Family Origins, Collective Property, and the State: Trends in Intergenerational Class Mobility"; and "Quotas and Careers: Trends in Intragenerational Mobility." Two distinct questions are addressed explicitly in each chapter: "Did socialism change the distribution of inequality?" and "Did socialism change the rules by which inequality was allocated and reproduced?" The questions are answered step by step with the use of unusually extensive mobility data taken from the "Social Mobility and Life History Survey," which was carried out in 1983 by the Hungarian Central Statistics Office on almost 25 thousand respondents. Chapter by chapter, the author makes her point that the socialist experiment in Hungary has led to significant changes in important distributions (i.e., educational, occupational, and class structure), but—in spite of radical interventions into allocation processes (e.g., "quota" systems controlling access to higher education and important occupational positions)—it failed in the attempt to change the rules of allocation (i.e., relative chances). Therefore, the chapter on education challenges the view that school reforms and "quotas" for children of different social background equalized relative chances for higher education. The chapter on trends in intergenerational mobility shows that the

socialist experiment did not lead to any significant and long-lasting equalization of relative chances of entry into the rapidly growing New Class. A similarly pessimistic picture is provided in the following chapter on career mobility, where the model of constant social fluidity proved to fit the data well enough to conclude that no significant change in relative chances for career mobility was generated by an evidently high inflow and class-based quotas controlling entry into the managerial class.

Unfortunately, only one chapter deals with postsocialist development ("Where Have All the Cadres Gone?"). Using data from the so called "Elite survey," carried out as part of the comparative project "Social Stratification in Eastern Europe after 1988," the author aims to decide between two competing hypotheses concerning the formation of new elites during the early stages of the transformation: namely the circulation and reproduction theories. The widespread and appealing thesis proposing that the old "nomenclature cadres" were able to convert their political capital into economic power and thus stay on track (the reproduction theory, or "political capitalism" thesis) is seriously challenged, although some support for the "political capitalism" thesis can be found in the data, particularly if we look at those "nomenclature cadres" who, at the time of the collapse of the communist regime, were relatively young and well educated. In any case, using Pareto's terminology, the author concludes that postcommunist Hungary has become one of the elite graveyards. It is also shown that entry into the new (postcommunist) elite was facilitated by a favorable social background (e.g., a well-educated father who was possibly a member of the old elite) and, of course, by a university education.

So far, only the book's strong points have been covered. There is, however, one weakness in that analyzing "the grand experiment in destratification," without looking at its effect on other areas besides mobility processes (i.e., income, poverty, wealth), especially those values and norms that actually keep a meritocratic stratification system running, is a fault with any project that aspires to deal with destratification. The fact that the author decides not to open these issues of destratification does not make the book any less interesting for students of social mobility. However, those who are interested in the wider social and psychological consequences of egalitarian experiments (e.g., the disintegration of meritocratic norms, values, concepts of life success, etc.) may feel that the book finishes before the interesting story really begins.

Where the World Ended: Re-unification and Identity in the German Borderland. By Daphne Berdahl. Berkeley and Los Angeles: University of California Press, 1999. Pp. 307. \$45.00 (cloth); \$16.95 (paper).

Andreas Glaeser
University of Chicago

Ethnographic accounts of Eastern Europe's transition from socialism to capitalism are living proof that ethnography (especially anthropological ethnography) has come a long way from the museological pursuit of cultural purity, which necessitated the careful mental undoing of the effects of rampant social change, to embracing social change as the very core of what it is trying to study. Since the possibility of the ethnographer's presence is very often owed to the fact of social change, this reorientation means that the presence of the ethnographer no longer needs to be seen as an embarrassment to be painstakingly edited out of a picture-perfect account of cultural purity. Instead, the ethnographer's presence can be used as a heuristic device in a reflexive movement. The particular form this interest in change takes is moreover informed by postmodern aesthetic sensibilities, in particular by postmodernist joy in change for its own sake, which is emphasized by artful ridicule of any hypostatization of beginnings or ends or unavoidable paths. From this perspective, then, change no longer has to be either seen as harbinger of a golden future or as the sure sign of a decaying golden past. Controlled reflexivity along with a nondidactic interest in change characterizes much of the ethnography of Eastern Europe's transition, where contemporary ethnographers' postmodern inclinations meet a population disenchanted and charmed with, both past and present, talking about state socialism and capitalist democracy with equally ambiguous mixes of satire and romance.

Daphne Berdahl's *Where the World Ended* is a perceptively researched, nuanced, and beautifully written specimen of the ethnography of Eastern Europe's transition from capitalism to socialism and of social change more generally. Theoretically, she chooses to study changing processes of identity construction in terms of local definitions and redefinitions of boundaries within larger political and socioeconomic contexts. As soon as the opportunity arrived with the unification of Germany, Berdahl began to work in Kella, a small village in former East Germany, which is literally a stone's throw away from the former Iron Curtain. Accordingly, the shifting interpretations and practices surrounding the east-west boundary before and after unification center Berdahl's account. She carefully reconstructs the significance of this boundary during GDR times, the effects of its closure, the villagers' attempts to make it more permeable, the fears and desires it instigated, and the everyday life within the confines of the GDR's border regime. She also observes firsthand the effects of the border's reopening and subsequent symbolic closure. In addition, she considers a wide range of other boundaries, geographical and symbolic: There is the historically significant boundary between the

Eichsfeld, the Catholic enclave to which Kella belongs and its Protestant surroundings, as well as the boundary between the various states that have at one point or another laid claim to the territory. There is the boundary between the villagers and the state, first the GDR and then the FRG, which Berdahl studies in terms of local acceptance and resistance to state regulations and national metanarratives. Neither acceptance nor resistance rarely ever amount to extraordinary gestures in either direction, but they are inscribed in the mundane practices and discourses of the everyday: during GDR times, for example, in the voluntary participation of border surveillance and defiant Sunday Mass attendance; after unification, in the whole-hearted participation in the consumer economy with the added twist of ostentatious preferences for formerly eastern brands, and ridicule during carnival of the new western regime and its failure to deliver promised or expected goods. There is also the internal division of the village into status hierarchies, which are both defined symbolically and more mundanely through the display and consumption of goods. Here, Berdahl once more demonstrates the usefulness of Bourdieu's capital theory in teasing out the systemic differences in the processes of status acquisition through consumption in the former GDR (participation in social networks assuring access to desirable goods) and now in the FRG (through participation in the labor market) and the ways in which villagers learn them, communicate them, and adapt to them. Finally, there are the boundaries between men and women and between people sympathetic to various ideologies and their respective other, between communists and Catholics in the GDR, and still between communists, Catholics, and happy consumers under the new western regime.

Berdahl is at her very best in describing how these different boundaries intersect and are combined to produce different meanings depending on the context in which they are evoked. Thus, for example, the geographical identity of the Eichsfeld informs a local understanding of Catholicism that is played out against the official church hierarchy; east-west differences in gender role perceptions get a surprisingly local spin; and a communist party member also attended Mass as an expression of local belonging, another quietly sponsored the erection of a cross, possibly for the same reason. By stressing the ambiguity and the fluidity of boundaries and therefore also of the identities they sustain, Berdahl thus falls in line with the contemporary critiques of more traditional understandings of identity construction processes, which depicted identity formation as proceeding in rigid grids of binary oppositions. With the help of well-chosen allegorizations of her claims in biographical vignettes or stories about tell-tale events, she empathetically paints a lively, warm picture of Kella and its residents through the transition from life in the GDR to life in the FRG.

Subnational Politics and Democratization in Mexico. Edited by Wayne A. Cornelius, Todd A. Eisenstadt, and Jane Hindley. La Jolla, Calif.: Center for U.S.-Mexican Studies, 1999. Pp. 376.

Jonathan Fox
University of California, Santa Cruz

This collection makes a major contribution to the comparative study of democratization. Conventional approaches to the study of regime transition differ over the relative causal weights assigned to elite conflict versus mass mobilization, to political parties versus social and civic movements, as well as to domestic versus international factors. What these frameworks often share, however, is the powerful assumption that the nation-state is the only relevant unit of analysis, especially for comparative research. For the growing number of regimes that fall into grey areas of partial and uneven political transitions, this assumption of national political homogeneity can obscure more than it reveals. Many regimes, which may appear to be democratic to those doing research from libraries or national capitals, in practice deny important segments of their citizenry full access to democratic freedoms. Analytically, the causal arrow between national and subnational political change can go both ways, and this volume begins to disentangle the complex threads that link a nation's center to its regions.

The open-ended outcome of Mexico's protracted process of political change, Wayne Cornelius and his colleagues argue, will depend to a large degree on the interaction between national and subnational arenas. Specifically, while national political competition has opened up dramatically in the past decade, and many important states and local governments have been won by the opposition, many other regions remain under entrenched authoritarian rule. Both analysts and state-level opposition leaders have predicted that Mexico's democratized states will drive the national regime transition ("the centripetal road to democracy" [p. 1]). The most recent evidence, Cornelius argues, suggests instead that "the subnational political arena will be the principal source of inertia and resistance to democratization, rather than the breeding ground for democratic advances. . . . So long as anti-democratic forces at the state and local levels can operate with relative impunity" (p. 11).

This collection provides strong support for this general hypothesis. Its balanced mix of case studies includes four studies of political party organizations, six studies of popular movements, and three essays that draw analytical lessons from center-periphery relations. Bruhn's study of the dozens of municipalities governed by the center-left party of the Democratic Revolution in the state of Michoacán concludes "that in centralized authoritarian regimes, alternation in power at the local level can serve more as a diversion than as a path to either local democratization or national regime change" (p. 44). She sees local progress as driven more by national change. Shirk then analyzes the center-right National Action

Party's (PAN) presence in and governance of the industrial city of León, Guanajuato, filling an important gap with his exploration of internal party dynamics. Espinoza's study shifts from the municipal level, assessing Mexico's first opposition party in power at the state level in Baja California Norte. The PAN governed there, in *de facto* partnership with then-President Salinas, led Espinoza to characterize the experience as "contained alternation" (p. 78). Calvillo explains a rare case of right-left opposition unity in the unsuccessful 1991 campaign for the governorship of San Luis Potosí.

Because of Mexico's opposition political parties' limited basis in civil society, this collection's broader focus on the social basis for democratic change is especially appropriate. Tavera-Fenollosa finds the roots of Mexico City's 1997 democratization in the unusually successful post-1985 movement of earthquake victims. Torres analyzes the internal dynamics of the broad-based, post-peso crisis debtors' movement. Hernandez Navarro explains the state of Oaxaca's 1995 indigenous rights reforms, which bypass political parties to legitimate widespread nonparty indigenous local governance practices. Rubin creatively explains the precedent-setting democratization of the Zapotec city of Juchitán through the interweaving of institutional and identity politics. Hindley shows how a regional Nahua movement emerged and successfully blocked a proposed hydroelectric dam in Guerrero. Harvey then provides critical context for the Chiapas rebellion with his analysis of previous indigenous movements for full citizenship rights. Mexico's indigenous peoples are the most excluded as well as the poorest social groups (at least 11% of the population, officially), so it is quite appropriate that four of the volume's chapters focus on their movements for democratization.

In the concluding section on center-periphery relations, Eisenstadt unravels how weakened presidential authority allowed authoritarian rule to flourish in the state of Tabasco (whose governor then became a leading presidential candidate by the fall of 1999). Snyder shows how neoliberalism's ostensible withdrawal of the national state from regulating key markets (in this case, coffee) creates space for diverse "reregulation" efforts by state governments. Prud'homme effectively compares Mexico's conflictive state-level elections during the 1988–94 period, most of which were resolved not by the voters, but through extrainstitutional pressures from both above and below. This collection persuasively reveals the limitations of most efforts to make nationwide, homogenizing generalizations about Mexico's regime in transition, with important implications for rethinking the study of regime change more generally.

Women and Democracy: Latin America and Central and Eastern Europe. Edited by Jane S. Jaquette and Sharon L. Wolchik. Baltimore: Johns Hopkins University Press, 1998. Pp. 250. \$49.00 (cloth); \$16.95 (paper).

Carol Mueller
Arizona State University West

Women and Democracy is a tightly woven anthology of case studies on women's mobilization under eight authoritarian regimes that continue through their democratic transitions and consolidations. Argentina, Brazil, and Chile provide the central focus for Latin America; Poland, Hungary, the Czech Republic, and Slovakia represent Eastern/Central Europe. Peru and Bulgaria are also included for comparison as countries in the two regions that are less developed both economically and politically. An introductory chapter by the editors and a conclusion by Philippe Schmitter complete the volume.

The close integration of the book chapters reflects the considerable organizational efforts of the two editors, who fostered the project through many stages. Much of the unity in the volume is found in the key questions that the editors ask of the material from the case studies in their introductory chapter. As scholars of comparative politics the editors' major concern is the absence of serious consideration to women's roles in the vast literature on the recent democratic transformations in Latin America and East/Central Europe. The present volume is designed to address this void by showcasing women's mobilization in the two regions. Jaquette and Wolchik analyze the case studies to determine the significance of women's roles in the transitions and how this participation affected the policies of concern to women in the new regimes as well as their forms of activism in posttransition democracies. They also try to determine what a gendered analysis could contribute to an understanding of prospects for democracy in the two regions. Finally, they explore the implications for feminist theory and theories of democratic transitions. Despite the excellence of the contributions and the depth of the editors' analyses, this is a great deal to ask of one short volume, and I was not disappointed or surprised to find that questions related to the nature of women's activism are better illuminated than those requiring an evaluation of the transition process.

Countries were selected because of the similarities of their locations in the "semiperiphery" and of a legacy of fragility and failed expectations in addition to their common history of authoritarian regimes. The two regions differ in the centrality of the role of the military in Latin America, in the recovery of national sovereignty from Soviet control in Eastern/Central Europe and, the editors argue, in lines of political division from left to right in Latin America and from *Gemeinschaft* to *Gesellschaft* in Eastern/Central Europe. Overall, the editors conclude that differences have outweighed similarities in women's political experiences in the two

regions: Latin American women mobilized around gender issues during and after the transition whereas gender, conceived as the quest for equality, had been preempted and discredited by the communists in Eastern and Central Europe.

As the editors indicate, previous experiences of bureaucratic-authoritarian regimes in Latin America have not precluded many strategies that are familiar to feminists in North America and Western Europe: the pursuit of electoral representation through voluntary and state-directed party quotas (Argentina, Brazil, and Chile), the capture of parts of the state in offices and ministries for women (Brazil's Council on the Status of Women and Chile's SERNAM), constitutional initiatives (Argentina, Brazil, and Peru), interest group activity, and attempts to heal divisions between women and progressive political parties (Argentina and Brazil). They have also created new institutions such as special police stations in Brazilian cities to aid women who are victims of rape or abuse, and active participation in international convocations of women beginning with the Mexico City celebration of International Women's Year in 1975. Although the solidarity and rich diversity of women's activism across race and class during the time of transition has given way, the editors argue that women are now recognized as "political actors."

Women's activism in Eastern/Central Europe has been quite different, at least in the short term. During the mobilizations that overthrew communism in 1989, women's participation was seldom framed in terms of gender; nor has gender been a major organizing principle in the transition and consolidation. The disappointment of Western European (and North American?) feminists with this state of women's activism in the former communist countries is directly addressed by at least one author. Writing about women's activism during 33 years of Kádárism in Hungary, Julia Szalai points out that it was women's flexible working hours based on their roles as mothers and caregivers for newborns, the sick, and the elderly that enabled them to develop the skills for running the secondary economy. Through the secondary economy that engaged three-quarters of adult Hungarians before 1989, families were able to develop new skills, pool meager resources, and build informal networks of information and exchange. Women's skills in negotiating, bargaining, and dealing with bureaucracies while taking care of family needs have also proved invaluable in the new economy. It is not immediately obvious to Szalai that women have much to gain by mobilizing in search of equality in education, labor force, and political participation, which existed, in theory, under communism. Hungarian women, like those in Poland and the Czech Republic have, however, mobilized to defend abortion rights when threatened by conservative legislators and to support services such as shelters for raped or abused women. There is also evidence that they will mobilize to defend the social welfare services they have lost or may lose in the transition to democratic, market economies. Despite this largely non-Western agenda, women's rates of political representation in Eastern/Central Europe are approaching those of Great Britain

and the United States, if not that of the Nordic countries and the Netherlands.

At the heart of the differences emphasized by the editors between women's mobilizations in Latin America and Eastern/Central Europe lies the political meaning of public and private boundaries in diverse types of authoritarian regimes and the persistence of these differences even as democratic transitions are consolidated. Even though women in Latin America have agendas now that are very similar to those of middle-class women in North America and Western Europe, some of the most compelling material on this region describes women's roles as "activist mothers"—the Mothers of the Plaza de Mayo during Argentina's "dirty war" and as organizers of communal kitchens in the shantytowns of Lima, which, at their peak in 1992, were serving 1.5 million meals a day. Such activism, based on the resources, opportunities, and grievances of women as mothers and family members, whether in Peru, Hungary, or Argentina, shows interesting parallels with that of low-income, working-class, and minority women in the United States described by Cheryl Gilkes, Karen Sacks, Sandra Morgen, Ann Bookman, Nancy Naples, and others.

Clearly, there is a great deal to be learned from the comparative study of how women mobilize when faced with diverse obstacles and opportunities. Under regimes of acute repression, where the public realm is dominated by the party or the military, the privacy of the family may be the only niche where there is sufficient trust and respect for the individual. Women's ingenuity in mobilizing under these conditions is remarkable. These patterns of activism teach their own lessons.

The Empowerment of Women in India: Grassroots Women's Networks and the State. By Sangeetha Purushothaman. Thousand Oaks, Calif.: Sage, 1998. Pp. 384. \$38.00.

Nancy A. Naples and Manisha Desai

University of California, Irvine, and Hobart and William Smith College

Sangeetha Purushothaman's primary goal is to help insert women's grassroots activism into social movement analyses—an endeavor that we, along with a whole generation of feminist scholars, have been engaged in for several decades. We read this book with great anticipation hoping for a rich case study located in the ever-expanding interdisciplinary and international social movements and feminist literature. Instead, we found a dissertation with its mandatory chapters on theory and literature (chap. 3–5) and a somewhat superficial analysis of data. The two chapters (3 and 7) that present data from the study are separated by several chapters that address different sets of literatures. As a result, the author frequently needs to repeat information. Throughout the book are very crude drawings or caricatures that have no value for the analysis.

The focus of the book is Swayam Shikshan Prayog (Self-Learning Ex-

periment [SSP]), a loose network joining 40 women's collectives and non-governmental organizations, which effectively gained the transfer of state resources to poor women in seven districts in Maharashtra. The network engages poor women from different parts of the state and the country in a learning process based on shared experiences and knowledge. Purushothaman outlines how the process of leadership development provided rural women with the skills to negotiate effectively for state resources. The emphasis on decentralized decision making and the development of community-based collectives, she argues, mirror the characteristics of Indian women's movement organizations more generally. She attempts to construct the work of this network as a social movement. This is the first point that needs further explication. When can an organizational form, regardless of structure, be considered a social movement? Purushothaman never addresses this question. While SSP represents the continuation of the 1960s and 1970s "non-party political formations" in India, which included the women's movements in India, Purushothaman demonstrates that SSP explicitly differentiated itself from it. Her case study might have been strengthened by exploring SSP as evidence for a more complex understanding of a plurality of women's movements and other social movements in India, rather than define SSP as the embodiment of an alternative movement for women in India.

After this introduction to the network, the author dialogues with the literatures on theories of social movements, community participation, and state theories in chapters 4–6, respectively. Her overview of the social movements literature is both simplistic and dated. Purushothaman is correct to point out that the dominant literature on social movements has been slow to incorporate the findings from feminist studies of women's political activism, which demonstrate the significance of women's networks and informal organizational forms. However, in recent years, feminist social movement scholars have successfully achieved the kind of synthesis Purushothaman seeks (see, e.g., Manisha Desai, "Informal Organizations as Agents of Change: Notes from the Contemporary Women's Movement in India," *Mobilization* 1 [2]: 159–74; Mary Fainsod Katzenstein, *Faithful and Fearless: Moving Feminist Protest inside the Church and Military* [Princeton University Press, 1998]; and Belinda Robnett, *How Long? How Long? African American Women in the Struggle for Civil Rights* [Oxford University Press, 1997]).

Furthermore, Purushothaman's lack of familiarity with feminist political theories, especially theories of the state, leads her to discuss the notion of engendering the state in rather narrow terms. For example, she summarizes feminist state theory as differentiating between "the benevolent father state," "the autocratic father state," and "the neutral state." Most feminist theorists who explore how the state is gendered hold a much more complicated view, one that highlights the contradictions of the state and the complex differentiation of state policies and practices.

While the theoretical framework circumscribes her ability to effectively analyze her case, she also fails to provide the richness of data that would

be needed for readers to assess her argument more effectively. Her discussion of key points such as how important the network was for women's learning and confidence is limited to information sharing about government programs and how to access state resources. There is no discussion of learning in terms of the women's sense of their personal or collective agency, of grassroots politics, and of other domains. In addition, the claims she makes from her ethnographic and interview data are often based on generic references to "some activists" or "other activists."

The conclusion attempts to highlight the significance of "decentralized, informal organizational forms, such as networks and women's collectives, in mobilizing resources for poor women and for facilitating their participation in developmental processes" (p. 306) without acknowledging similar claims made over a decade ago by contributors to *Women and the Politics of Empowerment* (edited by Ann Bookman and Sandra Morgen [Temple University Press, 1988]) and Gita Sen and Caren Grown (*Development, Crises, and Alternative Visions: Third World Women's Perspectives* [Monthly Review Press, 1987]) and in other relevant feminist literature (see, e.g., Chandra Talpade Mohanty, Ann Russo, and Lourdes Torres, eds., *Third World Women and the Politics of Feminism* [Indiana University Press, 1991]).

Our training in sociology rarely exposes us to the diverse interdisciplinary literatures that must be brought together for a study of women's activism and social movements. Therefore it is the task of the researcher to seek out the relevant knowledge base and theoretical frameworks. It is no longer possible, if ever it were, to rely on bounded disciplinary training to explore such complex topics as the relationship between gender, social movements, and the state. We recommend that you read this book in conjunction with the rich literature on women's community activism and economic development in different parts of the world. This will take you from sociology to political science to postcolonial studies to interdisciplinary feminist studies and beyond. But hopefully this journey will take you back to sociology with a richer lens through which to understand these important relationships.

Fields of Protest: Women's Movements in India. By Raka Ray. Minneapolis: University of Minnesota Press, 1999. Pp. 224. \$49.95 (cloth); \$19.95 (paper).

Verta Taylor
Ohio State University

This is one of the most important books about women's movements written in the past decade. Although feminist scholars proclaim the need for increased attention to women's struggle for equality around the globe, sociologists of the women's movement for the most part have failed to heed the call. In *Fields of Protest: Women's Movements in India*, Raka

Ray takes up the charge. The result is a rich and compelling account of the vitality and complexity of the contemporary Indian women's movement. The book has two intellectual projects. First, as a feminist scholar, Ray's objective is to highlight the long history of Indian women's collective efforts to challenge and change their subordinate position in the family, the workplace, politics, and other social institutions. Drawing from historical materials and interviews with activists, *Fields of Protest* presents a detailed and nuanced account of the Indian women's movement in two distinct locales—Bombay and Calcutta. The picture we get of women organizing on behalf of their rights contrasts sharply with popular accounts, found even in some feminist writings, that depict Indian women primarily as victims whose lives are restricted by gender oppression reflected in such torturous social practices as rape, battering, dowry murders, and widow burning. Ray's second intellectual project is also an interesting one: how to understand differences in the forms, identities, and strategies of the women's movement in different regions of India. It is this second objective that sets Ray's book apart from other writings on global feminism by subaltern scholars. Ray uses her comparative study of the distinctive feminisms in Bombay and Calcutta to identify political and cultural processes that are theoretically significant for understanding national, local, and regional variations in other social movements.

Ray's basic argument is that women's movements in India are neither homogeneous nor pure products of modernization and the diffusion of Western feminism. She provides a detailed analysis of feminist organizing in Bombay and Calcutta to demonstrate variation in the organization of local urban women's movements in India (both cities have politically affiliated and autonomous women's groups), the kinds of issues activists mobilize around, the success of particular ideologies and tactics, and, ultimately, the possibilities for feminist mobilization. The dissimilarities in Indian women's movements, Ray concludes, is not explainable on the basis of any real difference in women's objective lived experiences. The social and political circumstances of women are similar throughout different urban locales in India. Rather, they must be attributed to differences in the political fields in which the movements are embedded. In Calcutta, which is dominated by the Communist Party and a strong and traditional left culture, the women's movement functions more as a political party. Feminist groups use the party structure to address issues such as literacy, employment, and poverty, which do not directly threaten the gender status quo. The women's movement in Bombay, on the other hand, which exists in a more open, contested political field and culture, is more explicitly feminist and uses more autonomous forms of organizing to spotlight issues that are more threatening to men such as violence against women, religious fundamentalism, and women's restricted roles in the family.

As Ray herself acknowledges, we have long been aware of the fact that social movements do not emerge in a vacuum but emerge out of generalized political instability and protest. However, there are several ways in which the notion of political fields, which Ray develops in this study,

extends existing explanations of social movements. First Ray's definition of political fields directs our attention not only to the importance of the distribution of power by the state and political parties in an organization's environment, as postulated by political opportunity theorists. Drawing on Bourdieu's use of the concept of field, the notion of political fields also highlights the significance of the prevailing political culture, or the routine and legitimate ways of practicing politics in a given field, for understanding the emergence and distinctive character of a movement. The concept of political fields also brings greater conceptual clarity to the relationship between protest politics and conventional politics by directing our attention to the way protest movements are affected not simply by national politics, but by regional and local variations in the cultures, histories, and institutions of politics. A particular feminist strategy—for example, consciousness-raising geared toward the development of a feminist collective identity—might succeed in one field but fail dismally in another. Finally, by emphasizing the way women's collective interests are shaped by the local configuration of politics, the concept of political fields overcomes the excessive structuralism of some feminist analyses that link the emergence and collective interests of women's movements to modernization and demographic variables such as labor force participation, fertility, and education. To have accomplished all of this in such an eminently readable book is quite an achievement.

Gender and the South China Miracle: Two Worlds of Factory Women. By Ching Kwan Lee. Berkeley and Los Angeles: University of California Press, 1998. Pp. xiii+210. \$45.00 (cloth); \$12.95 (paper).

Diane L. Wolf
University of California, Davis

While economists ogled the East Asian "miracle" of the 1970s, feminist scholars documented the stark and harsh realities of the women workers who were at its basis, bearing much of the brunt of such rapid transformations. In Janet Salaff's book (*Working Daughters of Hong Kong* [Cambridge University Press, 1981]) on Hong Kong and in other works on Taiwan (Susan Greenhalgh, "Sexual Stratification: The Other Side of 'Growth with Equity' in East Asia," *Population and Development Review* 11 [2]: 265–314; Justin Niehoff, "The Villagers as Industrialists: Ideologies of Household Factories in Rural Taiwan," *Modern China* 13 [3]: 286–307), scholars consistency found that young factory women in East Asia were working hard at tedious jobs for periods up to and even beyond 10 years, remitting much if not all of their wages to their parents, and often ruining their health in the process. Contrary to most theoretical expectations, these young women did not gain more status or authority within their families due to their contributions to the family economy, and often remained painfully marginal within their families.

Ching Kwan Lee's work presents a much-needed sequel to this earlier research. Lee's book contributes a superb ethnography to the field of sociology, while updating feminist and sociological scholarship on contemporary women workers in South China. Most studies about women factory workers in Asia or Latin America focus on either one factory or multiple factories in one site. One very unusual aspect about Lee's study is the comparison between two factories belonging to the same firm but operating in two different sites—Hong Kong and Guangdong, China. Her study spans two very different state structures, work cultures, and populations of women workers. Lee's book successfully combines a feminist political-economic analysis with ethnographic data. Using the extended case method developed by Burawoy, she provides a sensitive view of factory workers' lives, concerns, and struggles, coupled with a vibrant analysis of labor markets and their resultant effects on factory work cultures in two very different settings.

The factory women she studied and worked alongside in Hong Kong were the "working daughters" in Salaff's study. Experienced factory laborers, these working-class, middle-age mothers are considered "matron workers," a status that brings with it a certain amount of respect from male supervisors and managers. Lee observed a flexible work culture that allowed workers to integrate their family lives with their work lives without stress. For example, they were allowed one to two hours of emergency leave without pay deductions for reasons such as needing to meet with their child's teacher. Furthermore, they could use the bathroom, call home, or receive private phone calls at any time. I found these examples extraordinary; they do not resemble anything I have seen in factories or read about in the literature on women and factory work.

Lee argues that the labor market in Hong Kong is controlled by familism, leading to a more flexible work culture she typifies as "familial hegemony." Moving from the Shenzhen factory site to the Hong Kong site, Lee felt as though she was on a passage "from hell to heaven." The Hong Kong factory was more pleasant physically, with a relaxed, "even playful," atmosphere (p. 138). Labor control was more covert and inconspicuous as workers had internalized the discipline of factory labor, which was expressed through a kind of self-regulated autonomy. She argues that a system based on familism led to a more humanizing labor regime that softened hierarchical relationships and forms of labor discipline.

By comparison, the "maiden workers" at Shenzhen's factory constituted a younger, less experienced migrant labor force. Although these young women worked for the same company as did the middle-age women in Hong Kong, the conditions differed vastly. They toiled under a much harsher and punitive regime based more on localistic connections and networks than on the familism Lee found in Hong Kong. In rural China, young women from poor rural families depend upon local networks to migrate and find jobs as they seek out the bright lights of city life. Like the families they are leaving, these networks are based on patriarchal relations; male relatives and locals from their native area become

their guardians in the city. The factory utilizes these localistic networks on the shop floor to ensure discipline and control.

Lee argues that the different ways in which the labor market is organized in each setting contributes to these different labor regimes. Furthermore, these labor market differences interact with gender dynamics to create different work cultures and work environments. Thus the familism in Hong Kong labor networks along with the gender dynamics and life cycle stage of the workers led to a much more tolerable labor regime than the labor networks and controlled work regime organized around locality found in Shenzhen. While Lee has crafted a compelling case for these arguments, one can well imagine the opposite occurrence—that work relations built on patriarchal family ties could be even more exploitative than those based on nativity. Indeed, historically, Western factories often utilized fathers to supervise and control their children's factory labor.

Lee's feminist theory of production politics builds on and adds a necessary corrective to Burawoy's theory of production politics. Lee spends a bit more time and emphasis than is necessary to challenge Burawoy's theory, but he was, after all, her doctoral advisor.

A native of Hong Kong who received her doctorate in the United States, Lee is both an insider to and an outsider in her fieldwork, to some extent. She is refreshingly honest about her position and earnest about the ethics of ethnographic work. She describes the field sites and her research dilemmas, such as being evicted from her research site, and includes her own voice. The insertion of the author's voice is another aspect that makes this book so unique since those adopting a political economy approach tend not to engage in much textual self-reflexivity, if at all. This book has already won the 1999 outstanding book award from the Asia and Asian American section of the American Sociological Association. It will be useful in courses on work, gender, East Asian studies, development, and fieldwork. I look forward to using it in my undergraduate gender and development course and in graduate courses as an exemplary piece of scholarship that creatively generates a successful cross-fertilization between sociological theory and ethnographic data.

Contesting Citizenship in Urban China: Peasant Migrants, the State, and the Logic of the Market. By Dorothy J. Solinger. Berkeley and Los Angeles: University of California Press, 1999. Pp. 463. \$19.95 (paper).

Zai Liang

City University of New York—Queens College

One of the major social transformations in China during the past two decades is the dramatic increase in internal migration, often referred to as the "floating population" or "tidal wave of migrant labor." They usually migrate from rural to urban China and do not have urban household registration (*hukou*). Even by conservative estimates, the total number

of China's floating population reached more than 70 million in 1997, arguably history's largest migrant flow in such a short period of time. In her new book, Dorothy Solinger, a renowned China scholar and a professor at the University of California at Irvine, provides the most comprehensive and systematic study of migration in China in the era of market transition. Drawing on Chinese journals and press, as well as interviews with government officials and migrants in several cities, Solinger pieces together a large story of societal transformation in the 1990s. The book is fresh in theoretical perspectives, solid in arguments, and rich in meticulous details concerning the lives and fortunes of migrants in urban China.

The book begins with a discussion of citizenship, markets, and state. Here, using the work of T. H. Marshall, Bryan Turner, and others, Solinger lays out the major argument of the book. She views the large volume of peasant migrants as an effort to contest over citizenship and makes a provocative theoretical argument concerning migrants in China. Solinger argues that unlike what is predicted by previous social theorists, the coming of market in China does not guarantee a full-fledged citizenship for peasants. To be sure, peasants now are able to migrate to cities but often without urban *hukou*, which has implications for employment, housing, and education of their children among other things. She contends that the outcome of peasant migration "must be understood in conjunction with institutional legacies left from the former socialist system" (p. 9). In fact, as Solinger argues, both the Communist state and the emerging market are stumbling blocks for migrants to become citizens. The state uses its rules and bureaucracies to prevent peasant migrants from becoming urban citizens. At the same time, the market encourages a mentality of profit making. Thus, we have the case of citizenship for sale in many cities in China. Of course, very few migrants are able to afford urban citizenship with such a high price.

In chapters 2–4, Solinger examines how three factors: state policies, urban bureaucracies, and urban rationing regime, affect the likelihood of migrating to cities and the life chances of migrants once in the cities. Using interviews with government officials and migrants, Solinger provides a balanced account both from the state and the migrants. Chapter 5 explains the rise of migration in China in the era of market transition. Three models of migration are examined, and merits and limitations of these models in the Chinese context are discussed: push and pull model, new-classical economic model of migration, and social network theory. Also notable in chapter 5 is Solinger's classification of two types of migrants: migrants from poor inland provinces who are more likely to be on their own and migrants from the coastal region who usually possess skills and craft. Chapters 6–7 explore the lives of migrants once they move into the cities. One of the most important findings in these pages is that two categories of migration (inland vs. coastal origin migrants) are sociologically meaningful because their life chances are indeed different. Perhaps surprisingly, these migrants from inland provinces, though with-

out much education and skills, end up doing better compared with migrants from coastal provinces who usually migrated with capital and skills. This is mainly because migrants from inland provinces are likely to be lured by the state recruiters to sign contracts with state-owned firms.

Compared with many quantitative studies of China's migrant population, which characterize most of the recent work in this area, a refreshing feature of this book is that we actually hear the voices of government officials and migrants themselves. Another important contribution is the author's creative use of the concept of migration networks. To be sure, the impacts of migration networks at the individual level are well known. Solinger suggests that labor contract recruiters add another layer to the migration networks, even though they are not necessarily mutually exclusive. What is also important for a book in China is that Solinger always keeps us in a comparative perspective. While examining migration issues in China, she brings experiences from other countries. Thus we learn which experiences are unique for China and which are universal across all societies.

I do have a couple of quibbles with the book, however. For example, Solinger's classification of inland versus coastal origin migrants, though it serves her analytical purpose, does have limitations. Solinger overlooks the socioeconomic variations within the coastal origin migrants. Take the well-known cases of migrants from Zhejiang, who operate clothing shops in Beijing. Clearly the life chances of shop owners are different from that of employees, even though both have skills and are from the coastal region. The successful Zhejiang entrepreneurs are able to buy urban *hukou* or afford the high tuition for their children, which are definitely not within the reach of an average worker in Zhejiang Village. Moreover, it may well be true that it is easy for migrants from inland provinces to get contracts from state firms, however, contracts come and go. Especially at a time when state-owned firms are laying off a large number of regular employees, migrant workers in state firms will be very vulnerable. In contrast, migrants from coastal regions with skills and craft as well as high likelihood of self-employment perhaps will be more likely to survive in China's changing and uncertain labor market.

From Education to Work: Cross-National Perspectives. Edited by Walter R. Heinz. New York: Cambridge University Press, 1999. Pp. 362. \$59.95.

Kathryn Borman
University of Southern Florida

Throughout the world, the transition from school to work is problematic for individuals who participate today in what the editor and the authors included in *From Education to Work: Cross-National Perspectives* refer to as the "international economy." Currently, individuals must negotiate demands for increasingly complex technical skills with inadequate links

between formal schooling and the workplace (with the exception of Germany, Japan, and the Scandinavian nations) and competition for positions above entry level from current employees in the firm. These are but a few of the challenges young people making the transition from school to work confront at every level—high school, the two-year college, and the university.

This volume argues that a developmental, life course perspective linked with an understanding of both the supply and demand sides of the equation captures the nuances and textures of the transitional process. Many authors in this volume go farther and argue persuasively for a regional analysis of the transition to understand how young workers must address specific local dimensions of the labor market. The central conclusion overall is not startling, namely that the institutional diversity characterizing transitions in Europe as opposed to North America is linked to the emphasis on the one hand (Europe) upon an “occupationally centered labor market” in contrast to an educational system (the United States and Canada) weakly linked to the labor market. We have heard it before, and, perhaps more discouragingly, its implications for policy change in the United States have been largely ignored. Young people in our society are still expected to experience a period of floundering, misaligned ambitions, and periods of underemployment (Barbara M. Schneider and David Stevenson, *The Ambitious Generation: America's Teenagers, Motivated but Directionless* [Yale University Press, 1999]).

The contributors to this book are fundamentally concerned with equity and opportunity in specific nations and take on the following central questions: Are there general trends in company restructuring and standard forms of work that restrict the life chances of the young generation? How do different societies respond to the decline of entry-level jobs and the upgrading of skill requirements? To what extent do they rely on market forces or attempt to build institutions that structure and support the transition from school to work for different groups? Other questions concern the role of education in preparing individuals for “jobs and careers . . . seen as proxies for a variety of potential careers that unfold their life-course consequences through the individual timing and sequencing of transitions” (p. 5). Taken as a set of comparative case studies, the collection of chapters places “young people’s experiences, orientations, and living conditions as they influence the timing and sequencing of transitions” (p. 5) in line with individuals’ opportunities and constraints. While understanding these and other structural dimensions is critically important for both individuals and policy makers, it is also necessary to understand the sense of agency individuals exercise in their orientations, strengths, desires, and decisions. This latter aspect is less well examined by contributors who are most concerned with institutional linkages and the gaps between them.

The book is divided into three major sections: Social Origin, Gender, and Transition Patterns; Education and Labor Markets; Work Experiences, Skills, and Credentials; and Changes in the Social Context of Tran-

sitions. A great strength of the volume as a whole is the reliance upon longitudinal studies from the four societies; two chapters use a case study approach, and two employ analyses of long-term societal trends influencing the school-to-work transition. The first section follows the editor's detailed introduction in which the volume's major themes—transitions and institutions in a life-course perspective and transitions and careers—are presented. The strongest chapters are those that utilize longitudinal data and connect findings to both useful analyses of critical policy concerns and key individual life course decisions. One notable example is the chapter by Thiessen and Looker "Diverse Directions: Young Adults' Multiple Transitions." This study followed male and female members of the 1971 birth cohort in three geographically diverse areas of Canada through a series of transitions from school to work. The startling finding was the extreme variability in pathways taken by the young people in their study with the sequencing of leaving the parental home, obtaining part-time or full-time work, and attending school bound up with marital and parenting decisions. Not surprisingly, these patterns varied for men and women with men's more "unidimensional" worlds emphasizing work and women's movement into adulthood more of a "balancing act." It will be interesting to follow the additional findings from this project.

The volume is a very important contribution to the literature on transitional experiences in the life course as well as to the sociology of education, the sociology of work, and research that seeks to link social structure with the lived experience of individuals at critical periods in their lives. It is also one of the first comparative collections to take the intersection of class, gender, nationality, and education seriously in analyses that chart the transition from school to work.

State-Making and Labor-Movements: France and the United States, 1876-1914. By Gerald Friedman. Ithaca, N.Y.: Cornell University Press, 1998. Pp. 331. \$55.00.

Kenneth H. Tucker, Jr.
Mount Holyoke College

In this skillfully argued and well-researched book, Friedman amasses a wealth of data on U.S. and French union membership and strike behavior in the late 19th and early 20th centuries. He compares revolutionary syndicalism in France with the business unionism of the American Federation of Labor. Friedman demonstrates that French radical unionism was as effective as its conservative U.S. counterpart in organizing workers. He critiques those authors who see workers as motivated by narrow economic interests.

Friedman contends that historians have focused on unions in isolation when explaining labor militancy, ignoring the interplay of union strate-

gies with the actions of employers and state officials. He states, "class formation is a dialectical process in which growing working-class militancy leads to a response by employers and state officials." (p. 217).

Friedman argues that variations in size, strike rates, and ideology between French and American unions were due to the differing political opportunities open to workers in each nation. In France, state officials protected union militants to some degree from employer repression. In the United States, unionists avoided radicalism out of fear of a hostile state allied with repressive employers.

Friedman traces these outcomes to the distinctive histories of France and the United States. After the defeat of slavery, U.S. republicans faced no conservative challengers to their rule. They allied with employers to repress an active and militant labor movement in the 1880s. In France, republicans confronted threats to their rule by powerful right-wing groups. They were forced to ally with labor to maintain their political power.

These differences emerged despite rich republican traditions in both countries. Republicanism united workers and capitalists against the monarchy in France, and this coalition opposed slavery in the United States. After the fall of slavery and the French monarchy, divisions between labor and capital undermined the republican consensus. Unionists argued that the republican principles of democracy and participation should be extended to the workplace; state officials supported the rights of owners and a separation of economic and political spheres, though the republican state in France supported workers right to organize and to strike. U.S. employers considered any regulation of the labor contract an attack on personal liberty, tying republicanism to a defense of economic individualism. French capitalists joined conservatives in opposing the Republic, encouraging the alliance of workers and republican officials.

This class hostility was demonstrated in the large number of strikes in the United States and France in the 19th century. In a mobile U.S. workforce divided by race and ethnicity, strikes often provided the venue for the mobilization of workers. However, they provoked employer and state reaction. In France, with its more homogeneous workforce, strikes did not give rise to the same extensive repression, because of the alliance of workers and the republican state. Strikes in the United States tended to be all or nothing affairs, as employers rejected collective bargaining and the state sided with them against labor. French strikes were more successful, often ending in compromises between workers and owners, as the state intervened to restrict the power of employers.

One of the most interesting aspects of Friedman's study is his defense of revolutionary syndicalism in France. Syndicalism is often dismissed as a small, ineffective, utopian movement. Friedman shows that revolutionary syndicalism was as viable a strategy for organizing workers as business unionism. Syndicalists were influential during strikes, when they provided examples of leadership and militancy that encouraged workers to join unions. They were particularly effective in organizing unskilled

and semiskilled workers in large firms. Syndicalism had been organized along industrial rather than craft lines since 1900, and it was able to integrate these workers into the union organization quickly. U.S. craft-based unions were ineffective in organizing these workers in the face of employer resistance.

Friedman concludes that many labor activists abandoned republicanism for Marxism, as they became disillusioned with the republican state. For Friedman, this was a mistake, as Marxism's economic determinism and view of the historical mission of the proletariat prevented labor from forging political alliances with other classes.

Friedman's otherwise excellent study is deficient in its view of culture and ideology. Despite his rejection of Orthodox Marxism, he sees ideas more or less mechanistically reflecting class position, as workers and employers were rational, calculating agents aware of their interests, and acted accordingly. Employers in particular seemed to be engaged in conspiracies against workers. Yet cultural factors that have a complex relationship to class struggle were at work in the decline of labor republicanism. By World War I, many workers joined employers and state officials in viewing the industrial order, economic growth, industrial efficiency, and increased productivity as inevitable and desirable, and downplayed workplace democracy. This acceptance of industrial society as a model of social organization was by no means inevitable, and calls for a nuanced understanding of the cultural role of science, technology, and labor in undermining the republican emphasis on democracy and participation.

Trade Conditions and Labor Rights: U.S. Initiatives, Dominican and Central American Responses. By Henry J. Frundt. Gainesville: University Press of Florida, 1998. Pp. 407. \$55.00.

Andrew Schrank
University of Wisconsin, Madison

The United States Trade and Tariff Act of 1984 allows Congress to deny "most-favored nation" (MFN) status to any country unable to prove that it is "taking steps" toward the defense of basic labor rights and the maintenance of adequate working conditions. While the question of China's MFN status has generated widespread controversy, the 1984 provision was first utilized in Latin America, and it has arguably proven most effective in the smaller, more vulnerable nations of the Caribbean Basin. The latter region therefore offers a particularly fertile environment in which to undertake an assessment of the relationship between U.S. trade conditions and Third World labor rights, and Henry Frundt is particularly well-suited to the task. After all, Frundt has not only established himself as one of the leading North American students of Caribbean Basin workers and working conditions, but he has also been actively involved in their struggles for well over two decades and therefore has access to a

wide variety of data, which would undoubtedly be unavailable to a less-experienced or committed investigator.

Frundt begins by asking whether the U.S. approach to the defense of Third World labor rights is effective. Will the denial, or potential denial, of preferential market access encourage otherwise recalcitrant governments to begin to adopt and enforce core labor standards? Frundt hopes to answer the question by using extensive data drawn from interviews with workers, managers, trade union representatives, human rights activists, and government officials; archival materials including petitions filed with the United States trade representative; and a thorough review of the secondary literature to examine the consequences of trade-based pressure in seven Middle American countries: Guatemala, El Salvador, Honduras, the Dominican Republic, Costa Rica, Nicaragua, and Panama. While he attributes the reform (or in the cases of Honduras and Panama, the enforcement and defense) of each country's labor code to U.S. pressure, Frundt is hardly sanguine. The new Salvadoran and Guatemalan laws have not given birth to widespread organizing campaigns or generalized improvements in living and working conditions, and the Honduran, Costa Rican, Panamanian, and Nicaraguan gains are in many ways ambiguous. The Dominican Republic offers a more "successful application of trade-based labor sanctions" (p. 227), but its experience is hardly typical. And Frundt's overall assessment of the U.S. approach is therefore best characterized as cautiously optimistic.

The enormity of Frundt's task—a comparative study involving primary research on seven different countries—must not be underestimated, and he has clearly written the definitive history of trade-based pressure on behalf of Caribbean Basin workers. While a number of the "victories" he cites (i.e., the contract approvals at Phillips-Van Heusen's Camisas Modernas plant in Guatemala and the Kimi factory in Honduras) have subsequently been compromised by plant closings and harassment, Frundt's fundamental belief in the efficacy of the U.S. approach is justified, for his detailed historical accounts offer numerous examples of government and corporate accession to the U.S. trade representative's demands.

Nevertheless, the book also reveals a widespread pattern of circumvention and backsliding, and it therefore fails to provide a convincing theory—as opposed to mere description—of the efficacy of trade-based labor standards. The limitation is in no small measure attributable to the author's research design, for each of his seven cases has been subject to U.S. pressure (p. 11), and he makes little systematic effort to measure the acknowledged variation in the degree of pressure or to control for omitted variables.

While he does assign each country a mark of "positive, negative, or unchanged" on each of nine "factors" (e.g., "basic rights," "contract approvals," etc.), and thereby provides a measure of his dependent variable (see table 13.2, p. 279), the latter effort is ultimately unsatisfactory. According to Frundt, "If we granted each of these factors equal weight, then

we would note eighteen improvements and thirteen declines, resulting in a slight improvement overall" (p. 278). But the factors should not be "granted equal weight," for the rights characterized as "basic" by the International Labour Organization are not only central to human dignity, but also provide the essential foundation upon which the other rights examined by Frundt rest—a point clearly established earlier in the book (see pp. 37–41 and pp. 60–61). Frundt's failure to acknowledge the distinction may account for the contradictions between his interpretation of table 13.2, which implies that El Salvador has performed better than Guatemala, Costa Rica, Honduras, and Nicaragua, and textual references that suggest the opposite (e.g., pp. 276–77).

In sum, the book offers more questions than answers, but given the importance of its subject matter and the breadth of its ambition, it is difficult to criticize. The manifold relationships between First World trade policy and Third World working conditions deserve ongoing scholarly attention, and Frundt's work will provide an essential point of departure for the many scholars who take up the task.

Forced Choices: Class, Community, and Worker Ownership. By Charles S. Varano. Albany: State University of New York Press, 1999. Pp. 409. \$24.95 (paper).

Vicki Smith
University of California, Davis

Charles Varano has undertaken an ambitious and important study of employee ownership at Weirton Steel, one of the most celebrated examples of a worker buyout in the contemporary United States. His research provides a much-needed glimpse into how employee ownership plays out in practice, how work reform is reshaping the industrial workplace, and how commitments to and ambivalence about these innovations are shaped by class relations but moreover by the dynamics of history and community.

Varano argues that the politics of the Weirton employee buyout can only be understood by looking at the historical currents preceding and shaping them. Relations between labor and capital at Weirton were forged over decades of both despotic and benevolent practices under E. T. Weir's system of paternalistic management. These practices culminated, by the early 1980s, in Weirton's workers being among the highest-paid steel workers in the world, as well as beneficiaries of generous health benefits and vacation policies. Over the years, the company had significantly invested in the town of Weirton, plowing money into the city infrastructure and sponsoring numerous community events and programs for its citizens. These gains, however, had been achieved after decades of company suppression of militant labor efforts and its endorsement of a company union, the Independent Steelworkers Union.

By the late 1970s, Weirton faced the conditions that threw corporation after U.S. corporation into crisis: competition from industries in other nations, the costs of not having adequately invested in new technologies, and, following Weir's death, governance by a distant parent company, National Steel Corporation, that lacked any sense of obligation to the community or workforce of Weirton. National viewed the Weirton mill simply as one operation among many that only mattered for its bottom line: adding to corporate profit. In 1982, National's disinterest came to shattering effect, when the corporation announced that it would no longer invest in the Weirton mill.

Weirton steelworkers were faced at this point with the option to "buy it or lose it," a forced choice to keep the mill alive by becoming employee owners. Varano documents how the buyout was designed and engineered by outside financial advisors and top-level managers, a top-down and undemocratic approach that immediately inspired the suspicion and opposition of many workers. Faced with the prospect of seeing the mill close, however, with the possibility of losing jobs and thus the very foundation of the community, workers consented, with considerable trepidation, to the employee stock option plan (ESOP).

Varano analyzes the course of events following the buyout and in so doing gives great depth to current debates about employee ownership. First, he fleshes out the significance of the fact that Weirton steelworkers became "workers by day, managers by night." Ownership, for the Weirton workers, was contradictory and frustrating. It was formal, based on equity through stock redistribution and profit sharing, but not substantive, falling far short of democratization in the workplace. In the initial stages, workers had no meaningful representation on the board of directors and were continually affronted by egregious displays of management power and control, as, for example, when managers received significant raises at the same time that steelworkers were making concessions in their wages. In general, the capitalist calculus of profit maximization and efficiency prevailed to the detriment of a workforce with elevated expectations about their role in the firm.

Deepening the indignity and inequality, employee empowerment on the shop floor ironically heightened class distinctions between workers and managers, continually reminding workers that they were subordinates in the workplace hierarchy. Varano argues that the employee involvement program further bureaucratized an already heavily bureaucratized mill, an argument that stands on its head the point of view held by some that involvement programs can hold out genuine transformative potential, both for workers and for hierarchical workplaces. Varano's reasoning seems quite sensible in the context of Weirton, in which corporate management consciously strove to delimit workers' participation in the global affairs of the company and thus were concerned to remind them that they were, after all, "just workers."

All the while, the Weirton workers' responses were shaped by their understanding of a moral economy. This understanding embodied a cri-

tique of impersonal power and arbitrary control, as well as an insistence on corporate responsibility, all based on the framework of mutual obligation constructed in the early paternalistic practices of E. T. Weir. Their outrage was fueled by their belief that National Steel was violating this moral economy. Their opposition to corporate policies continued up through the conclusion of Varano's research, finding life in periodic efforts to block board directives that would keep the balance of power in management's hands.

Forced Choices is a vivid reminder of the way that corporate managements often attempt to contain employee ownership and involvement and to manipulate ownership and involvement plans to increase profits rather than empower workers. Varano's thorough and fine-grained investigation of this case, made all the richer by his sustained immersion in Weirton for over a decade, makes an invaluable contribution to the meaning and implications of these issues.

Steelworker Alley: How Class Works in Youngstown. By Robert Bruno. Ithaca, N.Y.: Cornell University Press, 1999. Pp. x+222. \$45.00 (cloth); \$16.95 (paper).

Charles S. Varano
California State University, Sacramento

Qualitative research on the working class in the United States has a notable heritage. Ely Chinoy's *Automobile Workers and the American Dream* (Doubleday, 1955), Bennett Berger's *Working Class Suburb* (University of California Press, 1960), Herbert Gans's *The Urban Villagers* (Free Press, 1962), Richard Sennett and Jonathon Cobb's *The Hidden Injuries of Class* (Vintage Books, 1973), Lillian Rubin's *Worlds of Pain* (Basic Books, 1976), or David Halle's *America's Working Man* (University of Chicago Press, 1984) are a few of the most important studies in which ethnographic/interpretive data either advance social theory or settle (or raise) dust in intellectual controversies. In *Steelworker Alley* Robert Bruno discusses "how class works in Youngstown" in two ways: by placing it within a familiar debate over the embourgeoisement of the working class and by applying Ira Katznelson's theory of class formation from *Working-Class Formation: Nineteenth-Century Patterns in Western Europe and the United States* (Princeton University Press, 1986). Interviews and historical materials are used to answer whether these steelworkers became middle class, but one wonders if this question is as important today as it was in the 1960s when Berger and Gans (neither is cited) persuasively addressed it. An empirical application of Katznelson's theory would have been interesting had Bruno elaborated it sufficiently throughout his book. Unfortunately, while the richness of his account is welcome, the absence of clear links to social theory or contemporary disputes is disappointing. Students of social class will find his vivid portrait

of steelworkers' lives quite "thick," but they may also wonder "just where does it all lead?"

Bruno argues that steelworkers in Youngstown—including his father—did not become middle class but maintained working-class identities, loyalties, and consciousness throughout the post-World War II period up to the industry's bitter decline in the late 1970s. With numerous photographs complementing detailed description, Bruno walks us through neighborhood streets into the homes of family and friends to show how working-class lives were organized both in the workplace and in the community. Postwar "prosperity" was tenuous as steelworkers scratched, saved, and used credit to buy even the most basic consumer goods, signifying a middle-class lifestyle. Unfortunately, Bruno often provides far too much descriptive detail without clearly showing its relevance to whether the working class became middle class or not.

Likewise, Bruno begins a number of important arguments regarding class formation and consciousness in chapters on labor-management relations, but his analysis stops short of fruition. One would like to see how the conditions he describes either undermined or encouraged certain forms of class resistance. Specifically, why were efforts to stop the shutdown of Youngstown Sheet and Tube unsuccessful in mobilizing more support from workers like those Bruno interviewed? After reviewing the appendix of interview subjects, a stronger methodological strategy might have grouped the interviewees according to occupation, union local, and company in order to distinguish their views and actions regarding the shutdown. Instead, this listing of subjects is alphabetical rather than analytical, and one can only speculate as to these connections.

Bruno's discussion of stealing from the mill and sleeping on the job (pp. 118–30) is a fine example of how qualitative research can reveal informal and often illicit practices other methodologies overlook. These practices are reminiscent of the "indulgency patterns" characterizing labor-management relations of a gypsum mine studied by Alvin Gouldner in *Patterns of Industrial Bureaucracy* (Free Press, 1954). Bruno views stories of company theft as "working-class folktales full of normative messages" (p. 126) that reinforced class solidarity, but unlike Gouldner in *Wildcat Strike* (Antioch Press, 1954), he does not explain how this either undermined or supported class conscious resistance such as the wildcat strikes he discusses in the following chapter. Instead, sleeping and stealing, along with vulgar language and jokes are considered acts of resistance, but their political trajectory and structural sources remain obscure.

One frustrating aspect of this book is how Bruno moves between historical periods without delineating how contingent conditions influenced class formation. For example, when describing class solidarity within ethnically diverse communities, he combines information from 1975 with a footnoted reference from 1956 (p. 48). Mixing and matching historical periods is confusing, and it contributes to the analytical fragmentation that besets an otherwise richly documented account. Based on the ages of interviewees (none is younger than 59), this study is best described as

an oral history that might better address other questions than those asked by the author, such as how collective memory is organized, used, reinforced, and transferred to a generation of working-class descendants. Likewise, concise chapter introductions and summaries would have provided the reader with some theoretical and analytical direction. The most touching aspect of this book is the intimacy Bruno conveys, especially regarding his family and his childhood. One only wishes that his passion and compassion could have been accompanied by an equally satisfying analytical coherence. As it stands, this book will have to be read carefully for the interpretive gems that dot what is more a description than a sociology of class formation and consciousness.

Working Hard and Making Do: Surviving in Small Town America. By Margaret K. Nelson and Joan Smith. Berkeley and Los Angeles: University of California Press, 1999. Pp. 289. \$45.00 (cloth); \$16.95 (paper).

Ann R. Tickamyer
Ohio University

This book explores the complex survival strategies that working-class households pursue in the face of economic restructuring. It provides a detailed study of families in a rural area of Vermont that has had direct experience in the shift from manufacturing to service sector jobs, re-engineering, downgrading, downsizing, layoffs, and plant closures that mark the post-Fordist era of American industrial development. The authors scrutinize an intricate mix of multiple jobs and workers as they juggle formal and informal work, self-provisioning activities, and family and community commitments and exchanges. Their basic strategy is to compare the livelihood efforts of families with access to "good jobs" versus those stuck in unstable, poorly paid jobs that lack benefits or future prospects. Their findings run counter to both common sense and social scientific expectations: rather than serving as a substitute for stable labor force attachment, the ability to benefit from informal work and exchange; self-provisioning and multiple work; and worker strategies to improve income and living standards relies on the availability of at least one good job in the household. Those families who lack access to good jobs are also handicapped in their abilities to benefit from the myriad of other survival strategies that assist families to get by.

To arrive at these results, Nelson and Smith conducted in-depth interviews with family members selected through a snowball sample, backed up with a randomly selected telephone survey of county residents. In both methods, the analysis is restricted to two-parent households with at least one waged laborer and no barriers to similar participation in the formal labor force for the other adult. Families are divided into good job/bad job households, and virtually all subsequent analysis is conducted from this perspective. Detailed information is collected on a very large range

of livelihood strategies to determine who does them under what circumstances and with what outcomes.

Particularly valuable is the detailed account of different forms of informal and self-provisioning activities, making much finer distinctions than is often found in the literature. Numerous differences emerge in the abilities of good job households to deploy their labor and resources to augment income, community standing, and sense of worth compared to the stresses and uncertainties experienced by bad job respondents. Also of interest is the gender analysis that examines the ways gender shapes and is reconstructed by the multifaceted labor practices of these families. Not surprisingly, they find women's work is less valued and less under their own control, even when it provides greater income or security than the men's work. However, they do not provide as vivid an account of these practices as they do of the different work strategies, making this analysis more suggestive than definitive.

In conducting this study and reporting the results, the authors negotiate their way through fields that are land mined with academic controversies. The concepts of household, survival strategy, economic restructuring, informal economy, and the like that are the central subjects of this study are also sites of controversy for scholars interested in the complexities of late 20th-century work and family relations and the practices of piecing together livelihoods. They fearlessly maneuver their way through debates about the nature of informal labor, the relative utility of household versus family, and households as collective actors versus sites of oppression.

This need to deal with a contentious literature may partly explain the strategy adopted by the authors of reporting their findings in the body of the book, but providing a virtually parallel account in extensive and richly discursive endnotes that supply abundant scholarly sources as well as the opportunity to counter alternative interpretations and diffuse potential misinterpretations. Although the endnote form of documentation increasingly becomes a current fact of publication, in this case it is both annoying and gratifying. Annoying, since it is virtually impossible to pursue endnote threads while reading the text. On the other hand, the extensiveness and richness of these notes provide a resource of equal value to the narrative account.

Working Hard and Making Do is a dense text despite the generous use of quotes and profiles from interviews. This is a product of the detailed and finely drawn analysis that the authors bring to the complexities of the work and livelihood strategies they report. It is well worth the effort. The study adds to a growing body of research on the impacts of restructuring, the multiplicity of income generating and survival strategies, and the complexities of work and labor at the end of the 20th century. It is a welcome and valuable addition to research that recognizes that work is far more complicated, nuanced, and socially embedded than traditional labor market studies acknowledge.

Managing to Make It: Urban Families and Adolescent Success. By Frank F. Furstenberg, Jr., Thomas D. Cook, Jacquelynne Eccles, Glen H. Elder, Jr., and Arnold Sameroff. Chicago: University of Chicago Press, 1999. Pp. 318. \$32.95.

Joyce Ladner
Brookings Institution

What is life *really* like for families living in Philadelphia's inner city? How do parents in such an environment manage risks and opportunities for their adolescent children? In the absence of adequate income, how do they use social capital to maximize the advantages for their children's welfare and future success? These are some of the questions the authors explore in a superb book based on almost 500 interviews and extensive qualitative data collected over some five years.

The authors challenge the prevailing myths and misconceptions among some scholars and policy makers alike that often view such families as monolithic, disorganized, and unable to meet their most basic parental obligations. They found that poor families are as diverse in their composition and the way they manage their adolescent children's lives as those families with greater economic advantages. They take the reader beyond the traditional formulas and challenge the reader to look at these families from a more varied perspective. In this regard, they are following in a tradition that examined strengths of black families introduced by sociologists Andrew Billingsley, John Hill, Harriette McAdoo, and Robert Staples, among others.

At the center of their analysis is how the parents manage the scarce resources available to them in the rearing of their adolescent children. They describe family management as "a process by which parents build, invest, and deploy social capital—drawing upon social knowledge, information, and resources—in the interest of protecting and providing for their children and fostering their long-term prospects." The extent to which the families are successful in rearing high-functioning adolescents depend heavily on how the families use the resources at their disposal to manage their lives. We have grown accustomed to assuming that these families must manage risks and crises, but the authors say that they must also learn to manage opportunities. *The degree of resources these families and adolescents have and how they use them are critical determinants of their outcomes.*

They use the concept, social capital, to analyze social relationships that can be used as resources to achieve the parents' goals. Thus, social capital available to the subjects in this study were norms, reciprocal obligations, and opportunities for sharing information. The authors show that many of these families, although poor, possessed a reservoir of social capital that could, if managed properly, facilitate the successful adaptation of their adolescent children. Moreover, the greater the access to social capital, the more likely they are to succeed. Those who possess little social

capital or did not have an understanding of how to use this valuable resource are less likely to have favorable outcomes for their adolescents.

Parents use social capital to regulate their children's behavior, their environment, and their access to certain people, places, and things that can place them on a path to achieve success. These are the parents who are also more likely to understand the high correlation between behavior and outcomes, recognizing that unless certain behavior is managed carefully, the outcomes can produce lifelong consequences that greatly reduce the adolescent's ability to achieve. These indirect mechanisms of socialization, in which social capital is used, have important consequences for children's development and long-term welfare.

Despite their poverty and its attendant consequences, the authors found that most of the families in their study were "functioning reasonably well—some might even say very well considering that close to a quarter of the families were living in poverty and another quarter were experiencing chronic economic strain." They also provide convincing evidence that poor parents can be as skilled as the nonpoor, leading the authors to conclude, "we doubt that they differ greatly from parents outside the inner city."

A surprising finding was the lack of neighborhood advantage on such measures as academic performance, acting out, parents' assessments of their children's adjustment, or the children's self-assessment of their mental health, leading the authors to conclude that there was no neighborhood advantage linked to success in early adolescence. Neighborhoods are less capable of maintaining identifiable community-based norms. Advanced communication systems, access to jobs outside one's environment, and mobility diminished the distinctive advantages once held by neighborhoods. Again, good parenting exists in all neighborhoods and is not linked solely to those areas at greatest risk.

Parents' management styles depended on where they lived, their resources, personal experiences, perceptions of the risk in their environments, and how they grew up. It is not surprising that black parents were more restrictive and monitored their children's movements outside the home, regardless of class. Black children are more likely to encounter greater risks such as crime and discrimination outside the home.

This highly readable book breaks new ground. It should be read by those who are interested in learning more about families that may have little money but large amounts of social capital.

Chinese Christians in America: Conversion, Assimilation, and Adhesive Identities. By Fenggang Yang. University Park: Pennsylvania State University Press, 1999. Pp. x+238. \$50.00 (cloth); \$18.95 (paper).

Rebecca Kim and Min Zhou
University of California, Los Angeles

Chinese Christians made up less than 1% of the total population in China when Communists took power in 1949. During the Cultural Revolution (1966–76), all religious practices were banned, and Christian believers were persecuted. Chinese converts were publicly chastised with sarcasm—“one more Christian, one less Chinese.” Such a repressive past seems to be history. *Chinese Christians in America* denotes optimism. Fenggang Yang states that, since the 1980s, “hundreds and thousands of people” have turned to Christianity in the People’s Republic of China and that Christianity has now become the most practiced religion among immigrant Chinese in the United States. In the wake of such a renaissance of Chinese Christianity, Yang examines the process of conversion and religious practices of Chinese Christians in America and the ways in which they construct, reconstruct, and negotiate conflicting identities as Christian, American, and Chinese. Yang’s study is based on an in-depth ethnography of a Chinese Church in Washington D.C., where he is a member, and on participant observations and personal interviews with members in several other churches in the United States.

The central issue is whether immigrant Chinese can possibly become Christian and American while maintaining their identity as Chinese. Yang argues that religious conversion in the context of postmodern pluralism can be an act of preserving and revitalizing traditional cultures. Yang begins with a coherent review of theories on assimilation, ethnicity, and religion and offers a historical overview of Chinese emigration and ethnic identity reconstruction in the Chinese Diaspora. He then examines three identities of Chinese Christians in America in three separate chapters, explaining why and how Chinese immigrants become Christians (chap. 3), how they aspire to become American on their own terms (chap. 4), and how they negotiate their ethnic identity and preserve their cultural heritage (chap. 5). Evaluating the three different identities of Chinese Christians in America, Yang finds that, despite inherent conflicts and tensions, the church plays a dual function: while it naturally promotes conversion to Christianity and thus facilitates the Americanization of Chinese immigrants from diverse origins, the church unintentionally enables Chinese Christians to selectively preserve certain elements in the traditional Chinese culture and to strengthen ethnic solidarity around a common cultural heritage. Consequently, immigrant Chinese successfully negotiate and reconstruct their multiple identities as Christian Chinese Americans.

Yang’s book has significant implications for the theoretical understanding of assimilation and ethnicity. With increasing pluralism in post-

modern societies, Yang contends that assimilation becomes selective and immigrants become the agent, rather than the object, of change. Unlike the immigrants of the past who were expected to unilaterally assimilate into a presumably homogeneous culture, today's immigrants can negotiate their position in the host society and exercise control over the pace and course of their assimilation through religious conversion as well as reevaluation of their own culture of origin. As this study illustrates, converting to Christianity does not mean that immigrants indiscriminately leave all of their traditional cultural values behind and take on those of the host society. Rather, in the process of Christian conversion, immigrants, especially those who are seemingly assimilated and residentially integrated, are able to come together and collectively negotiate identity formation, and thereby increase ethnic solidarity. Particularly in a post-modern pluralistic society, Yang argues that traditional values can find solid support through the immigrant church. Yang's argument illuminates the study of immigrant adaptation and identity formation and encourages students of immigration and ethnicity to incorporate a multicultural perspective into their research on related topics.

Like other studies, Yang's book has limitations. As a book of "the sociology of religion," Yang seems to miss the opportunity to theorize his research findings and develop a model to explain the paradoxical outcomes of religious conversion and ethnic solidarity (Yang notes that his study is not theory driven [p. 10]). Another limitation is a lack of empirical data to substantiate the claim that there is a significant growth and increasing visibility of Chinese Christianity, an assumption on which Yang's work is based. Third, Yang deals with how Chinese immigrants "harmoniously" incorporate three seemingly conflictual identities—Christian, Chinese, and American—but he omits entirely the issue of race in his discussion. How does racial stratification affect their triple adhesive identity? Moreover, it remains unclear what "American" means to Chinese immigrants, how Chinese immigrants conceptualize being "American," and how the immigrant church influences the construction of an American identity. Fourth, Yang underestimates the latent social and economic functions of the immigrant church and argues instead that Chinese immigrants turn to religion for essentially religious reasons. Finally, Yang does not pay much attention to how socioeconomic standing affects religious conversion and identity negotiation. It is possible that Chinese immigrants of relatively high levels of education and income (who are Yang's primary research subjects) are more likely than their coethnics of lower socioeconomic status to have more options for religious conversion and more resources to successfully and harmoniously negotiate multiple identities.

Overall, Yang's book is a welcome addition to the literature on contemporary immigration and the sociology of religion. Those interested in how religion affects immigrant adaptation and identity construction in postmodern society should read *Chinese Christians in America*.

Culture on the Margins: The Black Spiritual and the Rise of American Cultural Interpretation. By Jon Cruz. Princeton, N.J.: Princeton University Press, 1999. Pp. x+288. \$49.50 (cloth); \$18.95 (paper).

Michael Rogin
University of California, Berkeley

Culture at the Margins brilliantly accomplishes two major tasks. By tracing the hermeneutics of the Negro spiritual from abolitionist weapon to object of cultural study, it also illuminates the origins of the cultural turn in social science, the methodology that has produced this book as one of its critical progeny. Thereby unravelled is a crucial strand in the history of how the white investment in black came to organize not only culture and politics in the United States but also social science.

Jon Cruz "attempts to track the trackers" (p. 203), he writes, rather than to listen to the music himself or engage in his own project of cultural retrieval, and by comparison with such studies as Lawrence Levine's *Black Culture and Black Consciousness* (Oxford, 1977) or Robert Cantwell's *When We Were Good* ([Harvard University Press, 1996] whose account of the folk music revival shares Cruz's focus on the trackers), his self-description is accurate. But this theoretically exigent and beautifully written account also turns on claims about the meaning and use of spirituals for the slaves. For the emergence and disappearance of the black subject is the hinge of the story Cruz has to tell.

From being first heard as barbarian instrumental noise accompanying forced labor, black music became the vehicle for what Cruz calls ethnosympathy, a guide to the inner lives of slaves. Frederick Douglass "almost single handedly placed [slave music] upon the analytical table" (p. 105), but before a small group of transcendentalist abolitionists could hear spirituals as protests not only against slavery but also against capitalist regimentation, slaves had first to sing them. They had to turn Christianity into an immanent critique of the society that had converted them. Before abolitionists discovered the black subject, slave subjectivity first emerged in social struggle, Cruz argues, through the religious frameworks built by early proselytization. The master's tool dismantled the master's house; the religious host (pun intended) allowed for "the viruslike spread of the subjectivizing process" (p. 80), nourished as well for abolitionist interpreters by Enlightenment individualism and romantic antimodernism. Although Cruz invokes Foucault and other theorists who link the subject to subjugation and normalization, the "historical eruption" of "new subjectivities" (p. 232) posits a subjectification-as-liberation that antifoundationalist and post-Enlightenment theorizing cannot escape.

In the wake of the Civil War, however—here the abolitionist commander of black troops Thomas Wentworth Higginson is the hinge figure—ethnosympathy shifted from the black subject to the cultural object, from testimonials (like the slave narrative) to artifacts. Ethnosympathetic pathos and the hunger for authenticity linked the abolitionist to the post-

Reconstruction interpreter of spirituals. But with former abolitionists turning into professional folklorists, the spiritual became not a protest against slavery, but an elementary form of primitive life to be classified and preserved as the culture that produced it was vanishing. Now romantic antimodernism coalesced with the birth of a taxonomic social science, humanitarian reform merged with managerialism, black culture was detached from interracial political and social relations and black lives, slavery and lynching disappeared, and the spiritual drove out all other black voices. With Reconstruction aborted, the spiritual came to express the imaginary relations of former slaves to the real conditions of black existence: "Go Down Moses," originally published as a cry for freedom and the first spiritual to appear in print, would reappear (though Cruz does not happen to mention this) in the mouths of faithful *Gone with the Wind* slaves marching to support their masters.

As black culture entered visibility, the black subject turned into an invisible woman or man. Now the Earl of Shaftesbury, an early white Negro, could hear the Fisk Jubilee Singers and wish he were black, for the spiritual was metamorphosing into "a critical-intellectual, pathos-driven parallel to minstrelsy" (p. 159). (Blacking up was the most popular form of mass culture in the 19th century). The stage has been set for Josh Kun to use *Culture on the Margins* in his October 13, 1999, San Francisco *Bay Guardian* music review to explain how the white midwesterner Nick Hexum (who leads the group 311) can sing "Come Original" in faux Jamaican patois.

A "cultural theft" model that counterposes black producers to white appropriators and consumers, black roots to white fruits, cannot account, Cruz shows, either for slaves forced to sing by their masters or for the genesis of the spirituals themselves. He offers a version of "cultural miscegenation" instead, but it is one far removed from melting pot, multicultural, or white Negro evacuations of power. "The cultural turn that opened up the chance for modern intellectuals to envision the deeper ties between social fate, cultural practices, and history also turned away from the opportunities of such broader envision," Cruz writes (p. 196). Showing how culture at the margins revitalized the center—alike in society and in social science—*Culture at the Margins* turns back in the right direction.

Refried Elvis: The Rise of the Mexican Counterculture. By Eric Zolov. Berkeley and Los Angeles: University of California Press, 1999. Pp. xiii+349. \$45.00 (cloth); \$18.95 (paper).

Pablo Vila
University of Texas—San Antonio

Eric Zolov has written a highly informative and well-documented book about Mexican rock music. Growing up in Argentina in the 1960s, I clearly remember rock singers such as Enrique Guzmán who sang Span-

ish versions of American rock songs. I also remember how Mexican rock suddenly disappeared from the scene late in the decade, and I always wondered why. I finally got my answer reading Zolov's impressive social history of Mexican rock.

Zolov is clearly aware of the ambiguous role most consumer artifacts have in processes of identity construction. Therefore he claims that, on the one hand, Mexican rock was an important actor in the process of undermining Mexican postrevolutionary national project, due to its subversion of the traditional age and gender roles that characterized the ideology of Mexican ruling party (PRI), "On the other hand, however, the new youth culture also appealed to many adults' perceptions of what it meant to be modern, to have access to global culture" (p. 8). Out of this ambiguity, the history of Mexican rock is characterized by both embracements and rejections by the Mexican state, the multinational record companies, and Mexican youths of diverse social condition.

The book has seven chapters, an introduction, and a conclusion. In the different chapters, Zolov follows the development of Mexican rock from the late 1950s to the mid-1970s, when Latin American folk-protest songs replaced rock as the music many young people used in their construction of identity. Thus, we have richly empirically supported chapters about the impact of an imported youth culture in the 1950s and the social conflicts that accompanied it; the containment of rock and roll via its Spanish-language domestication (that often meant the transformation of a social critique lyric into a romantic one), where "Spanish-language *rocanrol* came to embody the modernizing aspirations of a middle class in ascendancy, but stripped of the offensive gestures of defiance that defined the original" (p. 11); the minor role played by rock during the 1968 student movement; the development of a native Mexican hippie movement; the unfolding of *La Onda Chicana* in the early 1970s, because "anxious to align themselves with a global rock movement, on one hand, and to disassociate themselves with Mexico's earlier 'contained' rock success, on the other, [young Mexicans] sought to invent a musical and stylistic expression they could call their own" (p. 12); and the backlash of the counterculture after the Avándaro festival in 1971 (attended by more than 150,000 people). In a quite brief concluding chapter, Zolov continues "to follow the trajectory of native rock music as it reemerges from barrio culture to present itself as a frontal challenge to the status quo" (p. 13).

No book is perfect, and Zolov's is not an exception. While a highly informative and readable social history of Mexican rock, most cultural sociologists (in particular those working on popular music) will find the book a bit confusing in his treatment of the musical phenomena. On the one hand, Zolov's intellectual "others" are those scholars who propose the "cultural imperialism" thesis, and Zolov is quite skillful in showing the shortcomings of such an approach. On the other hand, Zolov continually hesitates throughout the book between two different, and highly problematic, theoretical explanations of why Mexican youth followed (or failed to follow) Mexican rock. Thus, in some parts of his book, Zolov

advances explanations linked to the “structural homology” theory (proposed by people like Dick Hebdige, Ian Chambers, Paul Willis, etc.), in which specific musical styles connect, necessarily, with specific social actors, and the connection is the product of a sort of “structural resonance” between social position, musical expression, and practice. In other words, *music “represents” people*. In other parts of the book, he aligns himself with the “musical interpellation” theoretical camp (advanced by Simon Frith, Richard Middleton, and others), according to which popular music performs the task of “offering” possible subject positions to its listeners, who eventually use that *music to “construct”* themselves as specific kind of *social actors*. I think that a more narrative-oriented approach could have helped Zolov to analyze his ethnographic data, due to the difficulties the homological argument has in explaining changes in the musical tastes of social actors who neither have changed their structural position in society, nor have modified the basic characteristics of their subculture; and for the problems the interpellation theory has in answering the question of why a musical interpellation is successful whereas others are not (Pablo Vila, “Interpellations, Narrative Identities, and Popular Music,” in *The Universe of Music: A History*, edited by Malena Kuss, [Schirmer Books/Macmillan, in press]).

All that said, I still highly recommend Zolov’s book for those sociologists interested in popular culture, rock music, youth culture, nationalism, Latin America popular culture, and processes of identity construction in general. It is a very good addition to the emerging field of music and identity studies.

SYMPOSIUM ON CLASS ANALYSIS

Toward a Sounder Basis for Class Analysis¹

Aage B. Sørensen
Harvard University

Satisfactory class concepts need to identify the mechanisms that produce the consequences of class membership, be they class conflicts or differences in lifestyles. Using a broad conception of property rights, the article proposes to base class concepts on personal wealth, that is, the assets a person controls. Two main class concepts are proposed: *class as life conditions*, based on a person's total wealth, and *class as exploitation*, based on a person's control over assets that produce economic rents. The former concept corresponds to empirical and Weberian class concepts, the latter to Marxist and neo-Marxist class concepts. The article shows that the class concept based on rent-producing assets accounts for recent developments in capitalism.

INTRODUCTION

There is an enormous literature on the concept of class that consists mostly of debates about which properties should be included in the concept. The result is a variety of class schemes and arguments that center around which class scheme is most appropriate for capturing the class structure of modern society. Dahrendorf (1959) argues that classes should be identified with authority relations. Ossowski (1963), and later Wright (1979), produces class schemes by cross-classifying property and authority or dominance relations. The class scheme identified with John Goldthorpe is based on property, employment, and authority relations (Goldthorpe 1987; Erikson and Goldthorpe 1992). Parkin (1979) and Murphy (1988)

¹ Earlier versions of this paper were presented at the ECSR conference, Rational Action Theory in Social Analysis: Applications and New Developments, Långholmen, Stockholm, October 16–20, 1997, and at lectures at the University of Oxford (November 1996) and Northwestern University (May 1997). I am indebted to the audiences at these lectures for helpful comments and to Hannah Brückner, John Goldthorpe, John Myles, Douglas Hibbs, Rolf Høijer, Christopher Jencks, Michèle Ollivier, John Scott, Annemette Sørensen, Ruy Teixeira, Erik O. Wright, and the *AJS* reviewers for comments, criticisms, and helpful suggestions.

emphasize relationships of closure, and Giddens (1973), the degree of "structuration."

The purpose of the original proposal for the concept of class, by Marx, is to explain inequality, social movements, social conflict, and political processes—to construct a theory of history. The mechanism that produces this extraordinary explanatory power is *exploitation* of the working class by the capitalist class, which produces *antagonistic interests*. Interests may be said to be antagonistic when the gain of one actor, or a set of actors, excludes others from gaining the same advantage. The incumbents of classes realize, through a process usually referred to as *class formation*, that they have these interests and form collective actors that engage in conflicts that eventually change the class structure and society.

Exploitation, for Marx and in this discussion, means that there is a causal connection between the advantage and the disadvantage of two classes. This causal connection creates latent antagonistic interests that, when acted upon as a result of the development of class consciousness, create class conflict. The causal connection also implies that the distribution of advantages or disadvantages can only be changed by changing the class structure.

The theory of exploitation as the cause of advantages and disadvantages among classes is a theory of inequality. It is a "structural" theory of inequality because the source of inequality resides in the relation between classes and not in the efforts and skills of the incumbents of these classes. In that sense classes are "empty places" as neo-Marxists like to say, using a formulation proposed by Simmel (1908), and the theory is a genuine sociological theory that can be contrasted to standard economic theory of how people obtain unequal returns on their skills, abilities, and physical assets in the market.

For most of this century, there has been agreement that the original conception of exploitation proposed by Marx is untenable. It is based on a labor theory of value abandoned long ago, even by Marxist economists. Since the labor theory of value is a point of departure for the whole theory, one should have expected that the formulation of a more adequate structural theory of inequality would have been a major concern for the revisions of the class concept. Nevertheless, the problem with the original exploitation theory has received very little attention, the main exception being the analysis of the exploitation concept proposed by the economist Roemer (1982). However, Roemer's very general conception of exploitation will not necessarily generate antagonistic interests that produce class struggles and revolutions.

The neglect of specifying an adequate theory of exploitation means to some that everybody has become Weberian (Murphy 1988). But, Weber's class concept proposes no structural theory of inequality that helps iden-

tify when class becomes relevant for social and political action. An essential ingredient of the original class concept as developed by Marx has therefore disappeared. This elision, of course, does not prevent the usefulness of class concepts used in empirical research to account for a variety of behavior and attitudes or social mobility, such as the concept proposed by Goldthorpe, or used to account for class formation processes, such as those proposed by the neo-Weberians.

The main contrast is not between a neo-Marxist and a neo-Weberian concept of class. A more useful distinction is between class as conflict groups where conflict originates in *exploitation*, and class as a determinant of individual actions and mentalities where these consequences originate in the *life conditions* associated with different classes. Both class concepts have properties that reflect the extent and type of resources or assets possessed by incumbents of class positions. My proposal sees class as based in property rights, as did Marx's, but the concept of property used here is broader than the legal property rights definition usually employed. It is a concept of economic property rights defined as the ability to receive the return on an asset, directly or indirectly through exchange (Barzel 1997). Some of these rights may be supported by the state, and they are then legal rights, but people also obtain advantages from rights that are not legally enforceable. Property rights define a person's wealth, and I suggest that the *class as life conditions* concept reflects a person's total wealth. Part of this wealth may be in assets that generate returns or payments that are *rents*. Rents are returns on assets that are in fixed supply because single owners of the asset to the market control the supply of those assets so that the supply will not respond to an increase in price. I propose to define *exploitation class* as structural locations that provide rights to rent-producing assets. Exploitation classes defined by the presence and absence of rent-producing assets have antagonistic interests because rents create advantages to owners of rent-producing assets, and these advantages are obtained at the expense of nonowners. Class locations defined by class as life conditions do not necessarily have antagonistic interests, because rent-producing assets may not be part of the wealth a person controls.

In the next section of the article, I briefly review the most important class concepts with an emphasis on the theories of inequality associated with these concepts. Next, I develop the two class concepts based on wealth. The last part of the article discusses the proposed class concepts' ability to account for recent developments.

THEORIES OF INEQUALITY AND CLASS CONCEPTS

Discussions of class concepts are often confusing because of the varieties of meaning of the term *class*. To clarify the discussion, it is useful to order

class concepts according to their level of theoretical ambition. At the bottom, so to speak, are purely *nominal classifications* of a population according to a dimension of stratification: for example, income, occupational prestige, or socioeconomic status. These concepts make no claim to the empirical existence of classes, identified with class boundaries, nor do they suggest why the dimensions of inequality, on which the classifications are based, come to exist. These concepts are nevertheless useful, despite what neo-Marxists sometimes argue, for describing differentials in all kinds of attitudes and behaviors.

At the next level of theoretical ambition, we find class concepts that make claims about the empirical existence of observable groupings with identifiable boundaries. I will refer to these concepts as *class as life conditions*. They may be detected by identifying different lifestyles associated with different living conditions in community studies (e.g., Warner, Meeker, and Bells 1949), or they can be approximated by a variety of class indicators such as occupation, education, income, sources of income, and residence, providing measures of the living conditions of different classes. These concepts are prominent in empirical research on classes and their consequences.

In recent research, the most prominent class scheme of this type probably is the class scheme proposed by Goldthorpe (1987) and elaborated in Erikson and Goldthorpe (1992). It has been widely used and found very useful in empirical research. Goldthorpe (1987) emphasizes that for a class to form, that is for collective class action, members of a class should at least have similar reactions to their class situation. This is partly a question of how similar the class situations are. Thus, a main task is to identify homogeneous class categories using occupational categories. The scheme is used to identify mobility patterns, and has also been used to analyze inequality of educational opportunity (e.g., Erikson and Jönsson 1996) and differentials in attitudes and behaviors (Marshall et al. 1988). It is often claimed to be a Weberian scheme, but Goldthorpe rejects this attribution. Grusky and Sørensen (1998) extend the approach of identifying homogeneous groupings to its ultimate conclusion arguing that unit occupations form the appropriate classificatory scheme. Indeed, if the concern is for identifying homogenous groupings to provide a useful site categorization for a variety of research purposes, this is a convincing argument.

Socialization and inoculation mechanisms are not specific to classes. The same mechanism would account for differences in attitudes and behaviors among persons raised in different local and national societies, or in different historical periods. Class as life conditions, therefore, is fundamentally a concept conveying the geography of social structure. These class schemes are descriptive of important differences between structural

locations, but they are not meant to predict revolutions. As with nominal concepts, this does not prevent them from being useful in investigations of differences in lifestyles which they are meant to capture. Recent formulation of class concepts has emphasized such cultural differences by sites (e.g., Bell 1987) and argued for new cleavages in postmodernist accounts (e.g., Eyerman 1994). A thorough review of these approaches is provided by Grusky and Sørensen (1998).

These class concepts do not propose or assume an explicit theory of inequality or how inequality produces interests, but presumably assume that the inequalities creating the different life conditions are created by the market or by some other mechanism. While Goldthorpe identifies market, employment relations, and authority as the bases for the scheme, the precise link between these defining relationships and the actual scheme is not specified.²

In theoretical discussions of the class concept, Weber is usually listed alongside Marx as the other main original contributor.³ Weber goes beyond descriptive concepts of class by explicitly locating class in the economic organization of society. "Class situation is ultimately market situation" (Weber 1946, p. 182). The need to realize and to preserve the full value of one's assets give rise to *economic interests* around which classes in conflict sometimes may form. Weber assumes standard economic theory of how people obtain unequal returns on their assets and resources. However, this theory does not identify under what circumstances economic interests will be antagonistic, resulting in conflict. It is perfectly possible that Weberian classes do not have antagonistic interests because one class obtains an advantage at the expense of another class. In perfectly competitive markets, with no transaction costs, there are no permanent advantages, or above-market returns, to be obtained at the expense of somebody

² Goldthorpe (1997) has recently begun this task, relying heavily on transaction cost economics.

³ The importance of the Weberian class concept in the literature on class analysis is a bit curious. In *Economy and Society*, Weber (1978) deals with class in two places, but both are very short fragments. While Marx can be said to never have given a single explicit development of the class concept, he certainly has class as the central concern of analysis in all of his writings. For Weber, there is neither a discussion nor an extensive analysis. Class simply seems not to have been an important concept for Weber. This is not for lack of alternative definitions and discussions of the concept proposed by Marx in Germany at the time when Weber wrote the fragments compiled in *Economy and Society*. Geiger (1932) lists 16 definitions, all by German language scholars. Except for Marx's definition, most are from the first decades of the 20th century. Since only Marx and Weber have been translated into English, Weber has become the main justification for developing class concepts that are alternative to Marx's, despite the fragmentary nature of Weber's writings about this and the lack of importance of class concepts in his writings.

else.⁴ Thus, class location is irrelevant. For economic interests to be in conflict, there must be advantages available that are not transitory.⁵

Weber does not emphasize this distinction between transitory and more enduring advantages that produce antagonistic interests. He provides two cues to what differentiates economic interests. One is the identification of economic interests with the goods and services persons sell in the market or the economic opportunities they face. The second cue can be inferred from the statement: "It is the most elementary fact that the way in which the disposition over material property is distributed . . . in itself creates specific life chances. According to the law of marginal utility this mode of distribution excludes the non-owners from competing for highly valued goods. . . . It monopolizes opportunities for profitable deals for all those who, provided with the goods, do not necessarily have to exchange them" (Weber 1946, p. 181). I will argue below that the idea of the importance of monopoly is relevant for class analysis,⁶ but for more specific reasons than the one suggested by Weber. The so-called neo-Weberians focus much attention on restrictions of access to classes, or closure, by conflating Weber's idea of status groups with class. The idea of class closure, emphasized by Giddens (1973), Parkin (1979), Murphy (1988), and others, suggests that classes have something to protect and want, but except for a general statement about property and credentials, there is no cue to when and if property and credentials give rise to antagonistic class interests forming a basis for class action.

At the highest level of theoretical ambition, we have the Marxist class concept, which provides a structural theory of inequality in the meaning I described above. The core process defining class relation, in Marx's class concept, is exploitation; that is, the process by which one class obtains an economic advantage at the expense of another class. In feudalism, the exploitation is transparent—the lords of the manor appropriate some of the product of the peasant, or, even more transparently, force the peasant to work for the estate for part of the work week without a wage. In capitalism, the exploitation is hidden, as the worker presumably voluntarily agrees to work for a wage. However, the wage does not reflect the value of the worker's product, which equals the labor power embodied in the

⁴ This is the standard result of neoclassical economics' perfectly competitive Walrasian model, where all profits and rents will be eliminated in equilibrium. Weber, of course, cannot be blamed for ignoring this idealized conception of the economy, but the failure of Weberians to identify structural locations providing significant advantages results in a weaker theory than Marx's class concept.

⁵ As pointed out by Hayek (1948), it is one of the ironies of a perfectly competitive market that there is no incentive for competition.

⁶ This point will be illustrated through the idea of monopoly rents discussed in the next section.

product—an abstract quantity not necessarily equal to the amount of labor embodied. The wage equals the exchange value or the price of labor that will reflect the cost of production of labor, as do other prices. The difference between the wage and the value produced is the source of the capitalist's surplus that generates profits, the end-all of all capitalist activity. The surplus belongs to the worker, and the capitalist therefore becomes rich at the expense of the worker. Clearly, the two classes should have antagonistic interests.

Marx's explanation of inequality and oppression is a very attractive one, as the history of Marxism shows. It is surely an alluring idea that the misery of the working class is caused by workers' spending part of their work on enriching the capitalist through the property arrangements of capitalism. It not only provides an explanation for inequality, it also points to an effective remedy: one must change the class relations that create exploitation. However, the worker's right to the surplus is a normative claim originated by Marx and developed in volume 1 of *Capital*. Surplus has no implications for observable economic quantities like prices. Marx realized this in volume 3 and argued that the sum of surpluses in labor values and the sum of prices will be the same. However, "as a general rule, profit and surplus value are really two different magnitudes. . . . The difference between profit value and surplus value . . . completely conceals the origin and the real nature of profit—not only for the capitalist, but also from the labourer" (Marx 1959, p. 58). This hidden nature of the source of exploitation makes it impossible to use empirically the theory and is the source of the difficulties of the labor theory of values encounters.

Exploitation is the appropriation of labor by the capitalist, just as labor was appropriated by the lord under feudalism. The distinction between the wage and the surplus value implies that the capitalist gains more the more surplus value he can get from the worker in a period of time. The *means of exploitation* available to the capitalist therefore are of paramount importance. Marx's class concept therefore acquires a dual dimension of legal ownership and domination, or power, that is seen as an essential element in the Marxist class concept in discussions and reformulations of the concept.⁷ Neo-Marxists usually distinguish a proper Marxist class concept from Weberian formulations by emphasizing the lack of attention to the means of exploitation in Weberian formulations. The preferred formulation is that class is defined at the *point of production*.⁸ Neo-Marxists

⁷ In addition to domination in employment, ideological and political structures also can be included, and we obtain the elaborate class concept developed by Poulantzas (1975).

⁸ The main exception is the class concept proposed by Roemer (1982) to be discussed below. For a critique that emphasizes exactly the need for a Marxist concept to have classes defined "at the point of production," see Wright (1982). Wright later revised his position (e.g., Wright 1997).

are right about Weber, but have focused on the wrong dimension of the Marxist concept, *domination*, to avoid the difficulties of the labor theory of value.

Two much-cited proposals for a reformulation of the Marxist class concept relies on the means of exploitation as a main element of the concept of class (Dahrendorf 1959; Wright 1979). Dahrendorf (1959) presents the most radical formulation of the Marxist class concept by eliminating the basis of exploitation, that is, legal property rights, from the class concept and keeping only dominance or authority.

The theoretical problem with Dahrendorf's reformulation is that he never details why authority relations should create antagonistic interests, the very root of class formation. Employment contracts are voluntary and represent an exchange of pay for the subordination to authority. In competitive labor markets, the possible discomfort and alienation felt by the subordination should be compensated by higher wages, as pointed out by Simon (1957) in his analysis of the employment contract. Therefore, there should be no antagonistic interests formed. Unless the labor theory of value is invoked, no exploitation is necessarily created by authority or domination.

Wright (1979, 1985) obtains class categories by cross-classifying ownership with authority in the manner also proposed by Ossowski (1963), with unusually clear justifications. Since most of the population has little or no property, most of the class differentiation is by authority. Wright's concepts have been widely used in empirical research. Only Goldthorpe's class concept, discussed earlier, has been equally popular in empirical class analysis. Wright (1979) claims that his first class scheme is based on exploitation theory, but never presents or discusses this theory. Later Wright (1985) adopts the exploitation theory proposed by Roemer (1982) and reformulates the class scheme accordingly, maintaining authority (now called "control of organizational assets"; Wright 1985, p. 79) as a main dimension of class relations.

Research using Wright's class scheme finds an authority effect on earnings, but an effect of authority on earnings does not require a class interpretation. Authority is measured as number of subordinates, and this quantity will clearly correlate highly with job ladders established in promotion schemes used in internal labor markets. To establish that the effect is a genuine authority effect, a differentiation between staff and line positions is needed, and this has never been done. It is difficult to justify an economic rationale for an income effect of authority per se—see Sørensen (1991) for further development of this argument.

Marx introduced the means of exploitation, in particular authority, as an essential element of his class concept not to explain the incomes of

managers and supervisors, but because the labor theory of value requires the dual dimensions of the class concept. If that theory is abandoned and replaced with neoclassical marginal productivity theory, the need for the means of exploitation disappears, for there is no distinction between exchange and surplus value in marginal productivity theory. In marginal productivity theory, the worker is paid what he contributes to the product: a lazy employee is paid less than a hard-working employee. Competition in the labor market guaranties that the capitalist pays no more and no less than the worker contributes to his production.

Marx did not employ marginal productivity theory, for he did not know about it. He certainly shared the belief in the competitive nature of labor markets under modern capitalism, and he may well have accepted the standard theory about these markets. As Roemer puts it, "The neoclassical model of the competitive economy is not a bad place for Marxists to start their study of idealized capitalism" (1982, p. 196).

Roemer attempts to formulate a theory of exploitation consistent with modern economic theory and Marx's original intent in developing his class concept, that is, that ownership of productive resources confers an advantage to the owner at the expense of the nonowner. Inequality in productive assets therefore creates exploitation: the value of what the poorer actor produces depends on the presence of the rich.⁹ Roemer gives a game-theoretical illustration of the idea by defining exploitation as existing if the disadvantaged group is better off by withdrawing from the economy with their share of the productive assets.

This concept of exploitation as created by unequal assets in a market economy has peculiarities. Roemer (1986) shows that if we let actors have different time preferences, then it is possible that exploitation status and wealth will be inversely correlated. Roemer's solution to this property of his exploitation concept is to propose abandoning the concept altogether.¹⁰ The possibility of formulating a satisfactory structural theory of inequality is therefore rejected.

The theoretically most ambitious concept, the concept of class as exploitation originating with Marx, does propose a mechanism of how antagonistic interests emerge and therefore how class conflict is generated. However, the theory rests on a labor theory of value that has been abandoned by economic theory. The various attempts to resurrect the concept by

⁹ This is a generalization of what sometimes is called differential or Ricardian rents, see Sørensen (1996) for a discussion.

¹⁰ According to Roemer, "exploitation theory is a domicile that we need no longer maintain: it has provided a home for raising a vigorous family, and now we must move on" (1986, p. 262).

invoking authority are unsatisfactory because it is not clear that authority is a source of exploitation and antagonistic interests. The proposal to see exploitation as grounded in inequality in all assets also produces unsatisfactory results.

There is another solution. This is to maintain Marx's insistence on property rights as the source of exploitation, but to not see all wealth as a source of exploitation. I propose instead to restrict exploitation to inequality generated by ownership or possession of *rent-producing assets*. Rent-producing assets or resources create inequalities where the advantage to the owner is obtained at the expense of nonowners. These nonowners would be better off if the rent-producing asset was redistributed or eliminated. A concept of *class as exploitation* based on the concept of rent is consistent with modern economic theory and therefore avoids the problems of the labor theory of value. It also avoids the anomalies discussed by Roemer.

CLASS AND WEALTH

Marx thought that classes were based on rights to the payments on wealth, and Weber thought property to be very important for the emergence of economic classes. Exploitation is a question of economic advantage obtained at the expense of someone else. The right to the returns on wealth is indeed essential for the distribution of these returns, I will show.

Rights to returns may reflect legal ownership. However, rights to the advantage provided by assets or resources need not be legal rights to be effective. Following Barzel (1997), economic property rights are properly seen as reflecting an individual's ability to consume a good or asset directly or consume through exchange, that is, to control the use of a good or an asset. Such economic rights may be enforceable by law and are then stronger, but they need not be supported by the state to be effective. Property rights are not absolute and constant, and they can be changed through individual or collective action to protect and enhance the rights. Such action incurs transaction costs that are the costs of transfer, capture, and protection of rights. Illustrations will be given below. When transaction costs are positive, rights are not perfectly delineated, and the transfer or protection of rights will be impeded or made impossible. Positive transaction costs may appear for a variety of reasons. Barzel (1997) emphasizes that some of the attributes of assets may be costly to measure and not fully known to actual or potential owners. These attributes are subject to capture by others who then obtain rights to the benefits from these attributes. Transfer of rights allowing an actor to realize the full value of his assets may be costly because mobility is costly or prevented by force. Col-

lective action needed to rearrange property rights that create a monopoly is costly.¹¹

For example, in the modern corporation, stockholders do not own all of the assets of the organization but share it with other parties inside and outside the organization that have rights to gains from various attributes of the assets. Managers obtain some gains because stockholders cannot fully control their use of assets because of lack of information. Other employees may obtain advantages (to be discussed below), for example, by retaining control over their effort. That ownership is divided does not mean that the concept of property as the basis for exploitation should be abandoned, as Dahrendorf (1959)¹² proposes. For example, the absence of individual legal property rights to productive assets in socialist society does not mean that individuals do not gain from controlling the use of an asset in such society. Only their property rights are more restricted in these societies, and it may be difficult to identify those who obtain the gains (Barzel 1997).

The broader concept of property rights proposed by Barzel implies that individuals—even slaves—usually have some property rights to assets under some circumstances (Barzel 1997, p. 105). This means that all individuals will have some wealth, even if it consists only of their ability to execute a task that can be exchanged for a wage.

A simple formalization may be helpful. Denote with v_j the value of resource or asset j , where value is given by the returns to j over the lifetime of the asset. These returns are usually monetary, but could also be social or psychological. Further, let c_{ij} be the right of actor i to asset j or the control i exercises over j . Then the total wealth possessed by i will be:

$$w_i = \sum_j c_{ij} v_j$$

where w_i is the wealth of actor i .¹³

¹¹ Barzel (1997) and others (see Eggertson [1990] for a review) emphasizing a property rights approach to the analysis of economic institutions see transaction costs as resulting from lack of full information and foresight. This reflects the focus on voluntary exchange. I suggest that actors may be prevented from realizing the full value of their assets also because of force or costs of combining action of several actors in collective action and maintaining such action.

¹² Dahrendorf (1959) rejects that property could be the basis for class formation. He bases the argument on the existence of inequality in state socialist society without private legal property rights to the means of production and on the emergence of the modern corporation with separation of legal ownership and control. Dahrendorf bases his argument on an overly restrictive concept of property, I hope to show.

¹³ This formulation is similar to Coleman's definition of power in a market exchange system (Coleman 1990). Coleman sees this formulation as the equilibrium outcome of the exchange process where actors exchange control over resources to maximize their interest in a system with no externalities. No such equilibrium conception is invoked

Individuals maximize their wealth by maximizing the return on their assets, employing them in the production of goods and services. This usually means that they will need the use of other assets controlled by other actors. Therefore, they need to transact with these other actors. A farmer needs land to maximize the return on the efforts and skills he may devote to farming; a worker needs an employer, machines, and raw materials, to realize the value of her main asset, that is, her labor power. These assets will often be controlled or owned by other actors—the landlord owns the land, the capitalist the machines. These assets can be bought by the actor needing them, or they can be rented. Rental here means transfer of use rights to the asset.¹⁴ The laborer can rent out her labor power to the capitalist in return for a wage, or the laborer can rent the capitalist in return for a profit to him. Such rentals are especially important with durable assets or resources, and even when assets are bought and owned, good accounting practice suggests calculating the payment of a rental to the owner.

The total wealth controlled by actors defines their class situation with respect to *class as life conditions*. The assets controlled will determine their incomes and the variability of their incomes. Workers will obtain wages as a result of their effort and skills, and their particular employment opportunities will be important for the variation in their earnings. Assets will be relevant for the respect and prestige received from the community when knowledge of these assets permits a collective evaluation of the standing of actors. The assets controlled will shape opportunities for transactions with other actors and therefore preferences or economic interests in the meaning of interests suggested by Weber. By shaping welfare and well-being, as well as economic opportunities and the investments that maximize these opportunities, the total wealth and its composition create the behavioral dispositions that are accountable for the inoculation and socialization mechanisms associated with class as life conditions, which I will amplify further below.

When individuals need to transact with other actors to get access to assets they need to realize a return on their wealth, the actors may be able to control the supply of the needed asset. Costs of mobility or other costs may prevent access to alternative suppliers, the supply may be inherently

here. Further, Coleman focuses on the exchange of any resource. The main interest here is in productive resources. For wealth in assets or resources to be valuable, the assets must generate a return and hence be involved in the production of something.

¹⁴ There is a possible confusion between rentals and rents. "Rental" refers to the transfer of use rights to an asset from one actor to another for a payment (a wage for labor, or interest for capital). These payments constitute returns to the holder of the benefit right to the asset. A component of this return may be an economic rent, to be discussed in detail below.

limited by nature, or the supplier of the needed access may have created a monopoly. This may allow actors controlling the needed asset to charge a payment for use of the asset that is greater than the payment needed to cover the costs of the asset. For example, the owner of a mine in an isolated location may gain an advantage from lower wages because workers are not able to find alternative employment.¹⁵ Workers are thus prevented from realizing the returns on their labor that they could have obtained elsewhere, and the mine owner has lower costs of production and therefore greater gains from production. The advantage thus gained from effectively being able to control the supply of assets is an economic rent.

Rents may also reflect lack of full information. Executives of organizations may obtain benefits far in excess of what is needed to secure their employment because they are able to control cash flows that stockholders are not able to monitor. Or the supply of the asset will be limited because its availability depends on the presence of specific other assets. In general, rents are advantages that prevent other actors from realizing the full return on their assets. Rents are crucial for the emergence of *exploitation classes* because those who benefit from rents have an interest in protecting their rights to the rent-producing assets, while those who are prevented from realizing the full return on their assets have an interest in eliminating the rents. Rents thus may create antagonistic interests and conflict.

To see how rents emerge, it is useful to consider the transactions involved in maximizing returns on productive assets more closely. The prices for the rentals of assets needed to maximize the returns on actor j 's resources are costs to j for those assets he does not own and returns to j for those assets he does own. This means his wealth is crucially dependent on the prices of assets relevant for him. These prices depend in the usual manner on supply and demand in the market. If the supply of a certain asset, for which actor j pays a rental, increases, the price will fall and actor j 's wealth increases since he has lower costs. Demand in a similar manner will influence the value of assets. This is the normal story.

Suppose now that actor k controls the supply of something j needs to employ her or his assets. Actor k may own land actor j needs access to in order for j to obtain a return on his labor and farming skills. They will negotiate a price for using the land and this price is a cost to j reducing the benefit he receives from his labor and skills. When negotiating the price, j and k will compare to what other farmers pay for land. In the long run, competition will ensure that a price emerges that will ensure j a sufficient revenue to keep him alive and able to work the land and compensate the landlord for whatever costs he covers, for example, for

¹⁵ The mine owner then has a monopsony.

fencing. Of course, j may try to buy the land instead of renting it, but for the eventual outcome this does not matter: the rental to the landowner is replaced by a rental of capital to finance the purchase, that is, of interest.

For the competitive equilibrium to occur it is important that the supply of land can vary in response to prices. It is also required that renters of land are mobile and thus able to take advantage of the rental offers provided by landowners. If these conditions are not met, if land is in fixed and limited supply or if renters of land—farmers or peasants—are prevented from being mobile by force or law, the owners of land can charge a rental price that is larger than the hypothetical competitive price that just covers their costs related to the ownership of land. The difference between the actual rental price and the competitive price is what is called an *economic rent*.

Rents are payments to assets that exceed the competitive price or the price sufficient to cover costs and therefore exceeding what is sufficient to bring about the employment of the asset. Thus, a rent on asset i can be defined as:

$$r_i = v_i^a - v_i^c$$

where v_i^a is the actual value of i and v_i^c is the value that would have emerged under competition and equal to the costs of making the asset available.¹⁶ These values are given by the stream of incomes generated by the asset over time.

The existence of rents depends on the ability of the owner of the asset to control the supply. I have already alluded to the classic example: the tenancy contract associated with feudalism. Part of the benefit from the land goes to payment for the labor of the peasant,¹⁷ and another part of the benefit goes to payment for capital expenditures on the land by the landlord. The rent benefit obtained from a tenancy arrangement is the remainder, that is, the payment not needed to employ the peasant and keep the land fertile. It is an advantage going to the landlord because of his rights to the returns on the asset that he controls.¹⁸ But the rent benefit

¹⁶ Rent is a component of what we ordinarily call *profit*, but profit as usually calculated includes a payment to capital, or interest, earned as payment for past savings and a component of wage to the owner of the asset for his management of the asset. The latter components are not part of the rent received when interest and wages equals market rates of return.

¹⁷ Classical economics saw land as the main source of rent, or ground rent to emphasize the dependency of the benefit on landownership (Marx, e.g., uses this terminology). Rent is "that portion of the produce of the earth which is paid for the original and indestructible powers of the soil" (Ricardo [1821], 1951, p. 67)

¹⁸ This is the typical arrangement. It can be argued that the benefit to the landlord is received in return for protection (North and Thomas 1973), but this can, at best, only account for the origin of the arrangement. Also Barzel (1997) argues that matters

forces a disadvantage on the peasant, since he does not realize the full value of his labor and skill.

The association of rents with land is not required. Rent will emerge on all productive assets that are in fixed supply and that actors need to maximize their wealth; or rent may be present as a result of transaction costs involved in getting access to needed assets. Alfred Marshall ([1920], 1949) devoted much attention to the concept of rent and generalized its applicability to benefits received from any productive resource or asset. He showed that rents also may appear as payments for the use of capital and labor in restricted supply; as payments for the use of unique combinations of capitals and labors, such as those created by certain technologies; and as payments for unusual and rare individual abilities that cannot be developed by training alone (musical talents, artistic creativity, athletic ability, etc.). Rents may be created in employment relationships when workers control their effort in an attempt to increase the advantage obtained by the wage because cost of monitoring prevents the employer from adjusting wages to effort. In general, the salient property is that a component of the payment obtained from the asset, or its return, is in excess of what is needed to bring out the supply of the asset.

The association of land with rent is not only an accident of history, it also reflects that tenancy arrangements usually are long term and rents therefore are long term. In a competitive economy, rents may emerge in industrial production as a result of an innovation or an import restriction. However, when others discover that there is an excess profit or rent available from owning a particular resource these others increase the supply of the resource if they can. This reduces the excess profit and eventually makes it disappear. Marshall (1949) calls such temporary rents *quasi rents*. These temporary rents are the typical rents in capitalist production and will become important in our discussion below.

For our analysis it is extremely important that rents are advantages to the owner of assets that are not needed to bring about the use of these assets. If the competitive payment is enough to make the landlord willing to let the farmer use his land, then any excess is in a basic sense unnecessary. It is an advantage costing nothing. The farmer has a clear interest in reducing, and if possible eliminating, the rent. The landowner has an equally clear and opposite interest in preserving the advantage provided by the rent. Rents therefore create antagonistic interests. Certain rents are especially important for social structure and social change. These are

may be more complicated dependent on the type of contract that exists between the peasant and the landlord. For example, under certain arrangements, the peasant may obtain advantages at the expense of the landlord, by depleting nutrients from the ground.

enduring rents that, resulting from enduring property rights to rent-producing assets, cause significant advantages and disadvantages. They are at the basis for class formation, as owners of such assets will protect their property rights to these assets, and nonowners will seek to eliminate these rights.

In summary, the individual's total wealth, as defined by her control of assets, will determine her life conditions and thus her class location in terms of class as life conditions. It will be argued below that the consequences of these conditions are not only dependent on the total wealth, but also on the overtime variability in the returns on that wealth (which define the variation in the value of the wealth). Part of the total wealth may generate benefits obtained at the expense of someone else, who would be better off with a different distribution of control or property rights to the various attributes of the assets. This *rent*-generating part defines class as exploitation. Below, I further develop these ideas, treating briefly first the idea of class as life conditions as total personal wealth and then treating in more detail the exploitation class concepts based on rights to returns from rent-producing wealth.

WEALTH AND CLASS AS LIFE CONDITIONS

As noted above, there is an abundance of research that shows that class as life conditions indeed is a powerful determinant of all kinds of outcomes.¹⁹ There is much less understanding of how these outcomes come about. We have, of course, a rich literature on socialization that demonstrates that class is associated with important socialization differences, we know about important value differences among different classes, and we also know about a host of lifestyles differences associated with different classes. However, this only moves the question one level back. What is it about the living conditions of different classes that accounts for these differences?

I propose that the answer is lifetime wealth and the expected variation in returns on that wealth for incumbents of different classes. There is

¹⁹ Also research using socioeconomic status as the independent variable provides abundant evidence. Socioeconomic status in the meaning of "goodness" (Goldthorpe and Hope 1974) seems to reflect people's belief about the living conditions associated with different occupations, and this is measured by the wealth of incumbents. There is no fundamental difference between what is measured by a class schema, such as Goldthorpe's schema (Goldthorpe 1987) and by socioeconomic status, except that the discrete class schema may capture nonvertical variation ignored by socioeconomic status measures. If socioeconomic status is grouped in discrete categories, we have a nominal concept of class as life conditions. There is some debate about whether discrete class schemes miss some socioeconomic effects (Hout and Hauser 1994).

abundant evidence that social class accounts for more outcomes the more homogeneous class categories are with respect to a variety of resources, or their wealth. It is important to consider not the cross-sectional distribution of income, but the long-term wealth profile that determines what economists call *permanent income and consumption patterns*. A person who obtains a higher education will orient her lifestyle not to the level of income in her youth, but to the long-term expected living conditions corresponding to the wealth associated with her human capital.

Further, the variation in the returns on the wealth is important, particularly for the socialization patterns that emerge in different classes. An older literature found strong differences between social classes in what was called "the ability to defer gratification" (see, e.g., Schneider and Lysgaard 1953). This literature was largely dismissed in the seventies because it was seen as reflecting an attempt to "blame the victim" (see Ryan 1971). More recently, psychologists and economists have suggested a different formulation of the same phenomena (see Ainslie 1992). People discount future rewards, often at very high rates. In particular, there are strong differences among social classes or different socioeconomic levels in time orientation, with persons at low socioeconomic levels having a much shorter time horizon than others. Those with high discount rates invest less in their health and education, and in the health and education of their children.

These differences among classes in time orientation or deferred gratification patterns reflect the level of uncertainty in living conditions or the variability of returns. Such uncertainty is not the fault of the "victim," but is a rational reaction to the expected high uncertainty of returns.²⁰ Banks also charge a higher interest rate with uncertain investments, and banks presumably are acting rationally. The impact of uncertainty on people's investments in themselves, and their children, should be greater, the lower the overall level of resources. Fewer resources give less of a buffer.

A person's total wealth has two main components. One part is personal, human, and physical wealth that is acquired mostly outside of the labor market in and from families and schools, but some is acquired from on-the-job training. The other component is wealth acquired from employment relations.

The personal part of wealth that exists, independent of the actor's employment relationships, has several components. The amount of human capital obtained through investments in training and health is particularly important. There may also be skills and abilities that command rents.

²⁰ There is recent evidence from a population survey (Dominitz and Manski 1997) that people's feeling of insecurity vary among population groups exactly as one would predict from the distribution of wealth and variability of wealth returns.

Finally, the amount of wealth obviously depends on the endowments of physical capital provided by the family of origin and augmented by the person through entrepreneurship and investment independently of his involvement with the labor market. This variation in endowments creates different incentives for investments in human capital and the like, and these differences explain the much emphasized nonvertical nature of the Goldthorpe class scheme, for example, the life conditions of farmers. This component of personal wealth is obviously of major importance for a full analysis of the class structure.

Individuals also obtain wealth from their employment relationships. They may have access to on-the-job training opportunities that increase their human capital. A component of the human capital acquired on the job may be specific to the job and the firm and give bargaining power to the worker. Because of specific human capital or collective action, the worker may gain above-market wages, increasing the value of his labor assets, and thus obtain a rent. The employment relationship will in these circumstances be closed in contrast to the open employment relationship characteristic of the competitive market. The resulting increase in the expected duration of employment relationships is crucially important for the variability of returns and therefore for the consequences of differences in wealth, because the shorter the employment relationships, the more variable will be the returns on wealth. The duration of the employment relation also is important for the amount of wealth obtained in the relationship. We should therefore expect persons in stable employment relationships to invest more in themselves and in their children. Professionals having large amounts of human capital and stable employment relationships should invest the most. Therefore, what Erikson and Goldthorpe (1992) calls the "service class" should be especially successful securing their children's future. Though no specific test of the idea proposed here exists, an abundance of research on social mobility and inequality of educational opportunity demonstrates the ability of professionals and others with high levels of human capital and enduring employment relations to secure the success of their offspring.

WEALTH, RENTS, AND EXPLOITATION

The issue for the formulation of a theory of exploitation is to define a process by which some holder of an economic property right obtains an advantage at the expense of persons without these rights. As shown above, wealth transfers made possible by the acquisition of rights to assets generating economic rents satisfy this requirement.

Rents satisfy the requirements of a structural theory of inequality. Rents are created by social relationships of ownership of rent-producing assets

(with the obvious exceptions of rents on natural abilities, to be treated later). Advantaged exploitation classes are positions in the social structure that allow individuals to gain control or economic property rights over assets or attributes of assets that generate rents; disadvantaged exploitation classes are defined by the absence of these rights.²¹ Changing the property relations that generate rents will change the distribution of wealth and hence the class structure.

The holder of a rent-producing asset has an interest in securing the continued flow of benefits, and those denied the benefit, a clear interest in obtaining the benefit by acquiring it or by destroying the social organizations that create the rents. When actors act on their interest, they create social organization and processes to protect or destroy rent benefits. These arrangements, well described by neo-Weberians, are processes of closure and usurpation (Parkin 1979) and the processes of moving from awareness of interests, through development of consciousness, to acting in pursuit of these interests (Giddens 1973). For the scenario to unfold, not only membership, but also interests must be enduring.

The distinction between temporary rents and enduring rents is very important for the analysis of class-formation processes. Class formation not only depends on stability of membership in structural locations providing antagonistic interests, as pointed out by Goldthorpe and Giddens. Class formation also depends on the rate of change in advantages or disadvantages provided by rents. This immediately suggests that structural conflict or class conflict should be more prevalent under feudalism than under capitalism, for rents are more permanent under feudalism. No revolution has occurred in an advanced capitalist society.

The importance of the distribution, over time, of the advantages provided by rents is often ignored. A cross-sectional inequality does not necessarily imply a longer-term advantage provided by an enduring rent. For example, according to human capital theory, the higher incomes of the higher educated compensate for higher training costs and do not create a permanent advantage over the lifetime of the person. Thus, skills acquired according to the mechanism proposed by human capital theory do not create rents and therefore not classes. This is generally ignored in the so-

²¹ Wright (1997) proposes a related definition of exploitation though it is not formulated in terms of the concept of rent. In addition to the causal link between advantages and disadvantages of classes, Wright requires that the advantaged class depend on the fruits of the labor of the disadvantaged class for exploitation to exist. Thus when the European settlers displaced Native Americans they did not exploit by obtaining an advantage at the expense of the Native Americans; they engaged in "nonexploitative economic oppression" (Wright 1997, p. 11). The European settlers clearly created antagonistic interests that brought about conflict, so it is not clear what is added by the requirement of transfer of the fruits of labor power.

called new class theory that sees classes emerge on the basis of skills and education (Gouldner 1979; Konrad and Szélenyi 1979). Education can, of course, create rents, but a measure of educational attainment as used, for example, in Wright's class scheme cannot separate the rent from the human capital component. The role of education in class analysis will be discussed further below.

The types of social organization and processes that emerge around rent-producing assets differ according to which type of asset is being considered. Feudalism can be described as an elaborate organization for the distribution of rent benefits based in land and mercantilism as an extension of the arrangements to cover industrial production. In modern industrial society, there are three main types of rents to be considered, already identified by Marshall (1949): (1) monopoly rents, based on monopolization of the supply of an asset, for example when a cable company gains a monopoly from a local government on distributing TV signals; (2) composite rents, formed by unique combinations of productive rents or asset specificity, for example when a worker has acquired skills only employable in a particular job; and (3) rents based on natural abilities and talents, for example, the height and ball-catching ability to make a professional basketball team. I will consider some of the main properties of each of these types.

Monopoly Rents

"Artificial" or social constraints on production create monopoly rents. The monopoly may have emerged "naturally" because of increasing returns to scale creating prohibitive costs of entering production for others, as in the production of automobiles. Often monopolies are created by governments as licenses or patents. Finally, social associations, such as trade unions or industry associations, who agree to regulate the production of something, create monopolies. In all these cases the supply of a product will not be sensitive to price, and rents will appear and persist unless the monopoly is broken.

For the sake of clarity, assume that the monopolist is working with production conditions that generate constant returns to scale so that average costs equal marginal costs. Nothing essential in the present argument depends on this assumption. Under perfect competition, output and price would be q_c and p_c and the price will correspond to the cost of the product. The monopolist is able to charge a price p_m above the price p_c that would prevail with perfect competition. This will cause an increase in revenue per unit produced that is an increase in the income of the producer over and above the amount needed to bring forth the production, that is, a

rent. In addition to the creation of the rent, and the corresponding increase in inequality, there will be a reduction in the wealth of society, for less is produced at the price p_m . This is the "deadweight loss" caused by monopoly rents and represents a welfare cost to society, which is a social waste of resources.

The increase in revenue to the monopolist is, of course, an advantage others might want. If others therefore successfully enter the market, the resulting competition might eventually erase the monopoly rent, lowering the price to p_c and increasing the quantity to q_c . When this happens, the temporary advantage to the initial producer is a quasi rent.

This scenario, of course, assumes that others can enter production. If there are prohibitive entry costs created by production technologies, governments, or trade associations, competition will be about obtaining the monopoly. Such competition is the typical case of *rent seeking*, that is, zero-sum competition over rent-producing assets. The efforts and other costs involved in trying to acquire the rent-producing property or resource of course reduces the benefit of the monopoly. Indeed, those who wish to acquire the monopoly should be willing to pay the equivalent of benefits to obtain it, so that the rent benefit completely disappears. The costs of rent seeking do not increase the production of society and therefore represent wasted resources (Tullock 1980). This waste is in addition to the waste represented by the deadweight loss.

The nature of the rent seeking depends on whether the monopoly can be traded in the market or not. If it can be traded, the sale may create a large transfer of wealth to those who obtained the monopoly first, and subsequent owners will not realize a rent. For example, it is often argued that rents received by farmers, as agricultural subsidies, produce higher land values and therefore higher interest payments, eliminating the initial advantage. Once established, the rent creating monopoly is difficult to eliminate, even when the monopoly is fully capitalized and the rent has disappeared. Clearly the new owners are strongly interested in receiving the rents they have paid for, even though the advantage to them has disappeared. An example, described by Tullock (1980), is a taxi medallion system, similar to the one in New York City. The medallions are sold, producing large gains to the initial owners, but only normal rates of returns to subsequent owners. Their existence creates a welfare loss to consumers. This loss can be reduced only by removing the restriction on taxi driving, something that is almost impossible to do without forcing losses on the present owners of medallions.

A variety of monopoly rents emerge in the labor market. *Employment rents* emerge when employment and jobs are closed to outsiders by the collective action of unions, by government-approved certification of pro-

fessions, and by other occupational licenses.²² Unions create rents when they close shops or ration employment through apprenticeship systems. Unions may also significantly alter the distribution of rents when they obtain egalitarian wage systems that will increase the wages to the least productive with the lowest market wages (for evidence, see Freeman and Medoff 1984). Professional associations create rents when they obtain certification limiting employment to the properly certified or when they gain control over the recruitment to the profession through control over educational institutions; medical schools are a good example of this. In general, educational credentials, used as rationing devices for employment or for access to employment-specific education for professions, create monopoly rents to those holding the credentials, a feature that I will discuss further below.

Employment rents create rent seeking as zero-sum competition for positional goods (Hirsch 1976) in what I have called "vacancy competition" (Sørensen 1983). Employment rents are not only monopoly rents. Positions may also be closed without the assistance of outside agencies such as unions or professional associations. In internal labor markets, closed positions can be created without collective agreements because of the existence of composite rents created by asset specificity, for example specific skills, to be discussed below.

The importance of monopoly rents is questioned by some. As noted, rents create a "deadweight" loss that is an externality reducing the welfare of society. There is an important objection to externalities in an idealized economic system, which is useful for the development of theory, presented by Coase (1960). Coase argues that, given an allocation of property rights, there will be no externalities, including those created by monopoly rents, if there are no transaction costs. One of his examples of an externality is cattle trampling on land, destroying corn. The cattle's owner is usually said to be liable for the costs imposed on the farmer growing corn, but Coase argues that this treatment is asymmetrical. The issue is whether the costs of avoiding the trampling is greater than the costs of fencing or of moving the cattle elsewhere. Rational actors will compare these costs and bargain about the cost of fencing in order to eliminate the externality. Applied to rents, this means that the nonowners of the rent-producing asset should negotiate a deal with the owner, to compensate him for the elimination of the monopoly. Thus institutions that exist over longer periods in a competitive economy, including those that appear to create monopoly rents, should be efficient and not rent creating, according to Coase. The gains to monopolies usually are smaller than the costs they impose on others. Therefore, Coase bargaining would eliminate the welfare loss

²² Bowles and Gintis (1990) uses the term employment rents to identify efficiency wages, i.e., above market wages created to induce effort, to be discussed below.

and with the abolition of the monopoly, output would increase to create the competitive situation. However, if rent-seeking costs have been incurred by the monopolist, his losses may be substantial so that he also will need to receive compensation for these costs. This may be difficult to achieve, as noted above. Therefore, when rent-seeking costs have been substantial, monopoly rents may persist, as has been pointed out by the rent-seeking literature (e.g., Tullock 1989).

Regardless of the problem with rent-seeking costs, there is a basic problem with the Coase argument when applied to larger categories of actors. The problem is argued by Dixit and Olson (1996). The individual rationality assumed by Coase, in formulating the idea of symmetric bargaining between two parties, does not necessarily create the collective rationality required when one of the parties, usually the disadvantaged part, is a larger group of actors. There are not only transaction costs involved with organization of a larger group to reduce free riding (Olson 1965). Dixit and Olson (1996) also show that even in the absence of transaction costs, the benefits of eliminating the externality per member of the larger group may be so small that no collective action will emerge. These organizational problems are what the class formation literature is about, or rather what it should be about. There is rich literature on social movements that address the problem of when interests will effectively be translated into action, emphasizing resource mobilization, political processes, and the collective action problem. This literature is curiously separated from the class formation literature developed by the neo-Weberians.

Composite Rents or Rents on Asset Specificity

When two separate assets or resources are so specific to each other that payment to their joint use exceeds the payment to each resource in separate use, composite rents emerge. Marshall's prime example of composite rents is the joint advantage to owners and employees of an advantageous market position (Marshall 1949). A specific example is the joint rent received when a mill is built on a water stream, to the owner of the mill and the owner of the stream. If there is only one site for the location of the mill, then the rents to the mill owner and to the owner of the water source cannot be separated: "There is nothing but 'higgling and bargaining' to settle how the excess of the value of the two together over that which the site has for other purposes shall go to the owner of the latter" (Marshall 1949, p. 520).

Composite rents emerge from what in the literature on transaction costs is called asset specificity. They emerge, for example, when workers have obtained specific on-the-job training and therefore are more productive in one firm than in another (Becker 1964). Also, monitoring and agency

problems may create composite rents. The composite rent creates a joint advantage that would disappear if the match between the firm and the worker were to be dissolved, so that employment relations become closed. There are two types of solutions to these problems.

First, the composite rent could be eliminated by organizing production so that the transaction cost problems disappear and employment relations become open. With respect to specific skills, this would imply eliminating the use of such skills. Such deskilling by eliminating the need for specific skills differs from the original deskilling idea made prominent by Braverman (1974), which suggests that capitalism will try to eliminate the need for all skills in the labor market. A general trend toward deskilling has never been established, despite many attempts, nor does it make theoretical sense that employers inevitably stand to gain by reducing the general level of skills required. However, reducing composite rents due to specific on-the-job training would be a plausible strategy.

The second solution is to reduce the importance of composite rents, without destroying closed employment relations to outsiders, by using organizational devices that increase effort. A large organizational literature on internal labor markets may be seen as analyzing organizational solutions to the problem of increasing the firm's share of composite rents. A prominent solution is the creation of promotion schemes to elicit effort. Promotion schemes capitalize on the interdependence of effort created by zero-sum competition over the wage and earnings differential provided by promotion ladders that are positional goods.

Promotion ladders create cross-sectional inequality. It is this inequality that is attributed to class by the Wright class scheme as an effect of authority. However, promotion ladders may be designed to provide less than the market wage at the start of the career, legitimized by training, and higher than market wages at the end of the career, according to the deferred payment theory (Lazear 1995). This pattern ties the worker to the firm and preserves the composite rent: she receives positive and negative rents depending on her seniority in the system. This does not mean that workers will obtain an overall surplus over their lifetime. As with investments in education, the cross-sectional distribution does not inform about the long-term advantage obtained. If the promotion ladders work as intended, they elicit effort and capture composite rents for the employer, quite contrary to the usual interpretation of the authority effect.

The interpretation of promotion systems as rent-capturing systems also implies that employers have an incentive to default on the positive rents the worker is to receive at the end of the career by dismissing the worker when these positive rents emerge. This will create "reputational" problems for individual firms, but if many firms collude in the practice, the reputational effects are diminished.

Another device used to elicit effort and to capture a larger share of the composite rent is the incentive wage system. By paying employees more than their market wage, firms increase effort since workers will be reluctant to shirk out of fear of losing their jobs.²³ Wright (1979) used such an efficiency wage explanation for the income advantage of "semiautonomous" employers.²⁴ In his latest class scheme, Wright proposes a similar explanation for the wage advantage of managers, called a "loyalty rent." The efficiency wage explanation is used by Krueger and Summers (1987) to account for the persistent wage differentials across industries that cannot be attributed to unmeasured worker characteristics or compensating differentials.

Rents on Natural and Cultural Endowments

Marshall (1949) suggests that rents emerge on "free gifts of nature" in the form of genetic endowments that result in the ability to produce something in demand. The rents directly reflect genetic endowment, as when genes are responsible for certain physical attributes facilitating certain tasks; for example, height for basketball players. Or, the rents obtain indirectly when an individual endowment facilitates training for certain skills, as in academic achievement. In the latter case, the endowment need not be genetic. Cultural endowments are important for learning, but hard to learn for those not socialized into a given culture or who lack the requisite cultural capital (Bourdieu 1977). Cultural capital thus may be seen as a source of rent similar to genetic endowment.

It may seem paradoxical to include rents on individual endowments as a source of structural inequality. However, these rents have important social consequences that connect to the class analysis literature.²⁵ In particular, rents on natural and cultural endowments have important consequences for the emergence of credentials.

General ability creates higher productivity and higher wages in many employments. Higher productivity may alternatively be obtained by training. The training costs needed by the less able create a surplus for the able,

²³ Bowles and Gintis (1990) see the creation of efficiency wage as the outcome of the "contested exchange" that defines the unequal power relations created in capitalism by unequal assets.

²⁴ Efficiency wage theory provides an explanation for involuntary unemployment (Solow 1979). The wage advantage makes unemployment a disciplining device, because the worker will often only be able to obtain the competitive wage after the layoff. We should therefore expect that layoffs are particularly frequent in industries with high concentration, such as automobiles and steel, consistent with evidence.

²⁵ For other social implications of rents on individual endowments, see Sørensen (1996).

assuming that equally productive able and untrained workers receive the same wage as the less able and the trained workers. With an expansion of demand for credentials—that is, an increasing demand for education—the rents become larger for the more able. They therefore seek even more education and higher and more expensive credentials. This self-stimulating demand is the main thesis of the credentialism literature (e.g., Collins 1979). The larger rents provide an incentive for institutions of higher education to increase tuition costs. They therefore share in the rents produced by credentials. Those rents then permit the hiring of prestigious faculty to train the easily trained, enhancing the reputation of these institutions and further increasing the rent on the credentials they confer.²⁶

Those possessing high credentials wish to secure an advantage to their offspring. This is facilitated by making cultural capital relevant for training. However, the very existence of credentials is also important. The superior ability of one's offspring cannot be secured, but much can be done to secure a valuable credential for the offspring by facilitating access to institutions providing valuable credentials. In the absence of such credentials, less able offspring from high-status backgrounds might have to compete for valued employment with the more able from more humble origins. The monopoly on employment ensured by the credential protects the less able from high-status background from being outcompeted by the more able from lower status backgrounds. Credentials thus increase the ability of high-status groups to confer their advantage to their less able offsprings and increase the advantage to their more able offspring. There are strong incentives for high-status groups to create credentials and closure as emphasized by the neo-Weberians (Parkin 1979; Murphy 1988).

In general, the differential rents generated by individual endowments imply that increasing equality of educational opportunity through educational expansion should increase the rents on natural and cultural endowments. Such policies therefore should be strongly supported by those who already obtain considerable rents on their endowments, such as professors.²⁷

EXPLOITATION CLASSES AND COLLECTIVE ACTION IN MODERN CAPITALISM

Rent-seeking activities create lobbies to influence the regulatory activities, subsidies, and welfare policies of governments. Social movements lobby to improve the welfare of the disadvantaged by granting them rents. Major

²⁶ I am indebted to an anonymous reviewer for this observation.

²⁷ Working-class parties in the past were indeed skeptical about policies to equalize educational opportunity (for Scandinavia, see Erikson and Jönsson 1996).

corporations lobby through campaign contributions designed to obtain the type of policies and regulations that increase the rents to these corporations. Rents divide owners of different productive assets—as when owners of land are in conflict with owners of industrial productive assets about corn tariffs—and they unite workers and capitalists to preserve import regulations and trade barriers that create rents to certain firms and industries.

That class action in modern capitalist society is about rent seeking and the protection of property rights to rent-producing assets clearly creates a different conception than what Marx had in mind when he analyzed capitalism in the late 19th century. Marx's conception, for example, does not require monopolies and asset specificities for the creation of advantage. The surplus created by labor will be a universal feature of capitalism, which will derive its nature from the relentless pursuits of ever-falling rates of profit. However, when the labor theory of value is abandoned, it is impossible to sustain the idea that there is a permanent "hidden" form for surplus in capitalist production in the manner conceived of by Karl Marx. The main class actions will be rent seeking, the protection of existing rents, and the destruction of rents.

It is an interesting question whether rent seeking, rent protection, and the destruction of rents might sustain Marx's grand scenario for the development of capitalism. Marx was certainly right about the dynamics of advanced capitalism. The engine of this dynamic is the pursuit of acquiring rent-producing assets through innovation and product development and by creating demand through advertising for profitable products. The relentless pursuit of advantages that exceed above-market returns—through the reorganization of firms and corporation, sometimes in the form of mergers and acquisitions, sometimes in the form of divestment—is also a pronounced feature of modern capitalism. These processes result in quasi rents that are usually quickly eliminated by competition. Individual capitalists gain and lose, and some obtain great fortunes. Even though their fortunes result from quasi rents, they are not destroyed by the elimination of these quasi rents. The process expands markets and produces globalism. The story is well known and well described.

Enduring rents to individual owners of capital require some type of collective action. The main form of collective action among capitalists is the establishment of cartels. Cartels may, of course, be hindered by government antitrust regulations. They may be effective despite such obstacles by various types of network organization among board of directors and the like, but the incentive to break an agreement is always present. A more effective strategy for obtaining enduring rents is to obtain help from the state to preserve an advantage: the granting of a license or some other form of protection from the entry of competitors. An army of lobby-

ists tries to obtain such advantages by informing legislators about the consequences and advantages of their actions. The rationale for state creation of monopoly rents is usually that some public benefit will be obtained by the regulation that otherwise will be lost: competent doctors, safe cars, and the family farm.

Can Marx's scenario for the class structure of advanced capitalism be sustained with the conception of rent-based classes as exploitation? Marx's emiseration prediction is usually taken as a main reason for revising his theory. Clearly the idea of increased absolute poverty of the working class, caused by increased exploitation, has been rejected by the economic growth that has occurred since Marx wrote. For a long period, also a decrease in relative inequality was observed in most societies. This decrease was replaced by an increase in inequality in the United States and many other advanced societies in the early 1980s, an increase that has continued since then. Nevertheless, it is not possible to sustain the idea that we find an increased polarization and homogenization of the working class. There is, however, substantial recent evidence that shows that capital has become very effective at eliminating the advantages of the working class in terms of rents obtained in the labor market. Eliminating these advantages has contributed to the increase in inequality.

Capital will gain by the destruction of monopoly rents in the labor market and by increasing its share of composite rents or destroying the source of these composite rents. The elimination of rents in the labor market benefits the capitalist when he or she benefits from the increased efficiency of production. He further benefits when his wealth is dependent on valuation of how efficiently he produces. The stock market provides this valuation. The stock market has, in the period where inequality has increased, very much increased the wealth of stockholders and rewarded rent elimination in the labor market. There are several main ways in which this has been achieved, and all have resulted in increased inequality: (1) eliminating rents created by collective action, in unions, (2) eliminating internal labor markets and composite rents, and (3) lowering the real value of the minimum wage.

Unions create rents in two ways. They may provide significant wage premiums for workers covered by union contracts. Nonunion workers may also obtain benefits when employers try to avoid unions. These benefits tend to accrue to workers who are highly skilled. As shown by Freeman and Medoff (1984), though unions do provide benefits, the rents are quite modest. The main effect of unions is to reduce wage inequality. Unions are especially effective at decreasing the wage spread between more and less productive workers. Unions may create substantial rents to low-skilled or otherwise less productive workers.

A well-known major change in the labor market has been the reduction

in union power. This is both a reduction in the number of workers that are union members and in the ability of unions to obtain wage increases and secure bargaining agreements. The reduction in membership has been from about one-third of the nonagricultural labor force to now 16%. The influence of the unions on the wage structure is far greater than its membership (Mitchell 1985). However, the decline in membership has also been accompanied by declining union power. The evidence is the increase in the number of concessions, the decline in the number of strikes, and the moderation of union demands (Mitchell 1985). For evidence on how these trends have contributed to the increase in inequality, see Fortin and Lemieux (1997). They also point to the importance of deregulation of highly concentrated industries, eliminating the composite rents obtained by workers and firms in these industries.

Closed employment and composite rents are widespread also in industries and firms without union presence, in internal labor markets, and for groups of workers traditionally not unionized, such as many white-collar groups. The composite rents obtained in these settings are eliminated by layoffs. Layoffs without recall reduce job security, but not necessarily employment. However, the loss of job security means also the loss of whatever rents the worker has obtained. With job security a worker can never do worse than his present job. If a better job comes along, he can move to this job, and the timing of this move need not have anything to do with increases in productivity. Therefore a system of closed employment, as in internal labor markets, produces career patterns that represent increases in rents only and not increases in productivity. These career structures are destroyed by downsizing.

Layoffs have increased overall over the last 15 years from 1.2–1.4 million jobs lost in 1979 and 1980 to 3.4 million in 1993. Layoffs grew to 2.62 million in 1982 and never fell below 2 million in the 1980s. They again increased in the 1991 recession and seem to have remained stable since. The proportion of white-collar workers in the total number layoffs has increased markedly to about 40% of the job eliminations (Bureau of Labor Statistics 1998). There appears to be a strong link between the occurrence of downsizing and the performance of company stocks, suggesting that the financial markets, in many cases, force downsizing on the firm (Love 1997).

Composite rents associated with internal labor markets can also be eliminated by job redesign and other changes in production technology. Or, they can be eliminated by removing asset specificity through outsourcing and subcontracting for labor. There is much talk about such changes, including how they could encourage the evolution of new types of employment relationships.

The elimination of employment rents through downsizing and job rede-

sign often means that workers are forced to look for new jobs in the labor market without much choice of which job to accept. This should mean that the next job after the downsizing is likely to be a worse job. It also means that the match between the downsized worker's productivity and wage is likely to differ from previous employment. There should be a closer match between actual individual productivity and wage level as a result of the job displacement. There is some evidence that suggests a tighter relationship between wages and productivity in the 1980s than in the 1970s (Levy and Murnane 1992; Mitchell 1985). Holzer (1990) thus finds a better match than do Medoff and Abraham (1981), but the two studies are not very comparable. Juhn, Murphy, and Pierce (1993) find the increase in inequality driven by increased returns to unmeasured skills.

Consistent with the idea of a stronger link between wages and personal endowments, we also observe a marked increase in within-occupation inequality. This is true for all occupations, but it is especially true for managers and sales personnel. In fact, for men the overall Gini coefficient rose from .315 to .332 between 1980 and 1989, but in the managerial and sales occupations combined, it increased from .322 to .353. For all other occupations, the Gini increased from .302 to .312 (Ryscavage and Henle 1990, p. 11). As inequality increased, structural locations seemed less relevant for explaining the variation in earnings.²⁸

The increase in inequality is very much driven by an increase in wages and earnings of the highest paid workers and stagnation or decline for others. The stagnation and decline follow from the rent destruction. The increases for the highest paid result from rent sharing with capital and may be legitimized by arguments that top managers were underpaid in the 1970s and therefore did not have enough of an incentive for doing their very best; in particular, they may have been more tolerant of rents to other employees (see, e.g., Jensen and Murphy 1990).

Stock market valuation clearly has been important for changing this situation. Another mechanism to increase manager incentive has come about through leveraged buyouts that force managers to squeeze all slack out of the firm to meet debt obligations. Leveraged buyouts also make top managers much wealthier. Finally, the increased competitiveness may have increased the rents on the abilities that boards of directors believe are needed in tough managers.

The declining real value of the minimum wage, until quite recently, also reduces employment rents for those less productive workers paid more than their competitive wage because of the minimum wage. This brings more poverty, for nothing guarantees that a competitive wage

²⁸ There is substantial evidence for Canada for the declining importance of "structural" or job characteristics as wage inequality increased (Myles, Picot, and Wannell 1988).

moves a worker above the poverty line. The rent destruction in the labor markets, except perhaps for the highest paid managers, leaves a labor market more flexible and more fluid, for fewer groups have anything to protect. The result is less structure, meaning less positional inequality, but more inequality overall. Thus, while greater homogenization overall may not have resulted from these recent trends, the destruction of rents in the labor market has created a labor market with fewer structural supports for the returns to labor. The idea of a homogenization of the working class can be sustained if it refers to the availability of structural advantages making earnings from work less dependent on individual endowments and more dependent on occupational choice and collective action.

Nothing guarantees that efficient labor markets create good lives. Rents are required in modern society to provide decent standards of living for the poorest part of the population. These rents are provided from the state in the form of income support and other welfare goods. The modern welfare state provides required support, but also creates an arena for rent seeking by all, including the middle strata with effective interest groups. It is beyond the scope of this discussion to deal with rent protection and rent seeking in the welfare state. Elsewhere (Sørensen 1998), I have provided a treatment of the breakdown of traditional norms around the provision of welfare goods and resulting increased rent seeking.

CONCLUSION

A sound basis for class concepts should be based on property rights to assets and resources that generate economic benefits. Property rights should be conceived of broadly. They are economic property rights defined as the ability to receive the return on an asset, directly or indirectly through exchange (Barzel 1997). Some of these rights may be supported by the state, and they are then legal rights, but people also obtain advantages from rights that are not legally enforceable. Property rights define an actor's wealth, and I suggest that the *class as life conditions* reflects a person's total wealth. Part of this wealth may be in assets that generate returns or payments that are rents. Rent distribution creates *exploitation classes* that may engage in collective action.

Class as life conditions is a very useful concept for analyses of how patterns of attitudes, behaviors, and socialization vary by location in social structure. A prominent example is the class concept proposed by Goldthorpe (1987; see also Erikson and Goldthorpe 1992). A major objective for constructing class schemes that account for different living conditions is to identify homogeneous groupings with respect to total wealth, type of wealth, and the variability of wealth over time. Such groupings will differ in the amount and type of investments they make in themselves and

their children. We therefore obtain class schemes that include nonvertical dimensions reflecting the type of wealth possessed and its variability over time, as generated, for example, by the stability of employment relationships.

The present proposal overcomes the evident problem associated with Weberian and neo-Weberian class analysis where there is no proposal for why anyone should be upset about their position in society and engage in class formation. Enduring rents identify antagonistic interests. Those who do not own a rent-producing asset suffer a disadvantage as the result of the rent. It is in their interest to eliminate the rent, and in the interest of the rent receiver to protect the advantage. The proposal I present here provides new insights. The concept of quasi rents suggests that monopoly rents often are transitory and their associated interests are therefore not enduring. Thus, not only stability of membership in structural locations and closure will be important for class formation, but variations over time in rent advantages are also important for predicting class formation. Rents provide a new interpretation of credentialism as a device to preserve and transmit advantages from one generation to the next with uncertainty about the ability of offspring.

The rent-based concept of class as exploitation provides an explanation for the recent increase in earnings inequality and for the practice of downsizing to destroy rents in the labor market.

The main class action will be by actors to seek rents, to protect rent privileges, and to destroy rents in structural locations, such as internal labor markets. The argument here implies that it is to the advantage of the capitalist class to produce a labor market conforming to the assumption of neoclassical economics, and I have tried to show that capitalism in the last decades has been successful in eliminating rents to labor. Eliminating rents in the labor market creates more efficient labor markets—that is, labor markets with less structure and more fluidity. A rents-free labor market will be one where simple class schemes are increasingly less applicable. The destruction of rents also creates more inequality within the labor market and produces more wealth that accrues to some of those owning means of production—for example, capitalists—whether they are old families, new entrepreneurs, pension funds, or graduate students with mutual funds. The resulting society conforms to Marx's predictions about the nature of advanced capitalism: "The bourgeoisie, whenever it has got the upper hand, has put an end to all feudal, patriarchal, idyllic relations. . . . It has resolved personal worth into exchange value, and in place of the numberless indefeasible chartered freedoms, has set up a single, unconscionable freedom—Free Trade" (Marx 1959, p. 323).

Thus, the main prediction about the development of capitalism from

rent-based class theory is that rents will disappear from structural locations in the labor market. This will result in a structureless society, without the nooks and crannies of social structure we have come to expect because feudalism is slow to disappear. The result is the transfer of wealth to those who have rights to rent-producing assets, even though these assets usually are quasi rents, for the wealth created by quasi rents is not destroyed when the rent is destroyed. As a result, we see increasing wealth inequality (Wolf 1995).

If Marx's grand scenario for advanced capitalism is interpreted as having to do with the distribution of rents, it is sustained. Rent seeking creates the dynamics of capitalism, and the destruction of rents in the labor market creates a structurally more homogeneous working class, that is, a working class without structural supports for its welfare.

REFERENCES

- Ainslie, George. 1992. *Piconomics*. New York: Cambridge University Press.
- Barzel, Yoram. 1997. *Economic Analysis of Property Rights*, 2d ed. New York: Cambridge University Press.
- Becker, Gary S. 1964. *Human Capital*. New York: National Bureau of Economic Research.
- Bell, Daniel. 1987. "The New Class: A Muddled Concept." Pp. 455–68 in *Structured Social Inequality*, edited by Celia S. Heller. New York: Macmillan.
- Bourdieu, Pierre. 1977. *Reproduction in Education, Society and Culture*. Beverly Hills, Calif.: Sage.
- Bowles, Samuel, and Herbert Gintis. 1990. "Contested Exchange: New Microfoundations for the Political Economy of Capitalism." *Politics and Society* 18:165–222.
- Braverman, Harry. 1974. *Labor and Monopoly Capital*. New York: Monthly Review Press.
- Bureau of Labor Statistics. 1998. "Worker Displacement, 1995–98." Labor Forces Statistics from the Current Population Survey. News release, Aug. 19, 1998. Washington, D.C.: U.S. Department of Labor.
- Coase, Ronald H. 1960. "The Problem of Social Cost." *Journal of Law and Economics* 3:1–44.
- Coleman, James S. 1990. *The Foundations of Social Theory*. Cambridge, Mass.: Harvard University Press, Belknap Press.
- Collins, Randall. 1979. *The Credential Society: An Historical Sociology of Education and Stratification*. New York: Academic Press.
- Dahrendorf, Ralph. 1959. *Class and Class Conflict in Industrial Society*. Stanford, Calif.: Stanford University Press.
- Dixit, Avishnah, and Mancur Olson. 1996. "The Coase Theorem Is False: Coase's Insight Is Nonetheless Mainly Right." Unpublished paper. University of Maryland, College Park.
- Dominitz, Jeff, and Charles F. Manski. 1997. "Perceptions of Economic Insecurity: Evidence from the Survey of Economic Expectations." *Public Opinion Quarterly*. 61:261–87.
- Eggertson, Thráinn. 1990. *Economic Behavior and Institutions*. Cambridge: Cambridge University Press.

- Eggertson, Thráinn, and Jan O. Jönsson. 1996. *Can Education Be Equalized? The Swedish Case in Comparative Perspective*. Boulder, Colo.: Westview Press.
- Erikson, Robert, and John H. Goldthorpe. 1992. *The Constant Flux*. New York: Clarendon Press.
- Eyerman, Ron. 1994. "Modernity and Social Movements." Pp. 707–10 in *Social Stratification: Class, Race and Gender*, edited by David Grusky. Boulder, Colo.: Westview.
- Fortin, Nicole M., and Thomas Lemieux. 1997. "Institutional Changes and Rising Wage Inequality: Is There a Linkage?" *Journal of Economic Perspectives* 11, no. 2 (spring): 75–96.
- Freeman, Richard, and James L. Medoff. 1984. *What Do Unions Do?* New York: Basic Books.
- Geiger, Theodor J. 1932. *Die Soziale Schichtung des Deutschen Volkes: Soziographischer Versuch auf statistischer Grundlage*. Stuttgart: F. Enke.
- Giddens, Anthony. 1973. *The Class Structure of the Advanced Society*. London: Hutchinson.
- Goldthorpe, John. 1987. *Social Mobility and Class Structure in Modern Britain*, 2d ed. Oxford: Clarendon Press.
- . 1997. "Social Class and the Differentiation of Employment Contracts." Paper presented at the European Consortium for Social Research conference, Rational Action Theories in Social Analysis: Applications and New Developments. Långholm, Stockholm, October 16–18.
- Goldthorpe, John, and Keith Hope. 1974. *The Social Grading of Occupations: A New Approach and Scale*. Oxford: Clarendon Press.
- Gordon, Milton M. 1958. *Social Class in American Sociology*. Durham, N.C.: Duke University Press.
- Gouldner, Alvin. 1979. *The Future of Intellectuals and the Rise of the New Class*. New York: Seabury Press.
- Grusky, David, and Jesper B. Sørensen. 1998. "Can Class Analysis Be Salvaged?" *American Journal of Sociology* 103:1187–1234.
- Hayek, F. A. 1948. "The Meaning of Competition." In *Individualism and Social Order*. Chicago: University of Chicago Press.
- Hirsch, Fred. 1976. *The Social Limits to Growth*. Cambridge, Mass.: Harvard University Press.
- Holzer, Harry J. 1990. "The Determinants of Employee Productivity and Earnings." *Industrial Relations* 29:403–22.
- Hout, Michael, and Robert M. Hauser. 1992. "Symmetry and Hierarchy in Social Mobility: A Methodological Analysis of the CASMIN Model of Class Mobility." *European Sociological Review* 8:239–66.
- Jensen, Michael C., and Kevin J. Murphy. 1990. "Performance Pay and Top-Management Incentives." *Journal of Political Economy* 98:225–65.
- Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skills." *Journal of Political Economy* 101:410–42.
- Kalleberg, Arne L., and Ivar Berg. 1987. *Work and Industry: Structures, Markets and Processes*. New York: Plenum Press.
- Konrad, Gyorgy, and Ivan Szélenyi. 1979. *The Intellectuals on the Road to Class Power*. New York: Harcourt Brace Jovanovich.
- Krueger, Alan B., and Lawrence H. Summers. 1987. "Reflections on the Inter-Industry Wage Structure." Pp. 17–47 in *Unemployment and the Structure of Labor Markets*, edited by Kevin Lang and Jonathan S. Leonard. Oxford: Basil and Blackwell.
- Lazear, Edward P. 1995. *Personnel Economics*. Cambridge, Mass.: MIT Press.
- Levy, Frank, and Richard J. Murnane. 1992. "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations." *Journal of Economic Literature* 30:1333–81.

- Love, Geoff. 1997. "The Diffusion of Downsizing among Large U.S. Firms, 1977-95: The Role of Firm Status." Ph.D. dissertation. Harvard University, Joint Program in Organizational Behavior, Harvard Business School and Departments of Sociology and Psychology.
- Marshall, Alfred. (1920) 1949. *Principles of Economics*, 8th ed. London: Macmillan.
- Marshall, Gordon, David Rose, Howard Newby, and Carolyn Vogler. 1988. *Social Class in Modern Britain*. London: Unwin Hyman.
- Marx, Karl. (1848) 1959a. "The Communist Manifesto." Pp. 315-55 in *Capital, the Communist Manifesto, and Other Writings by Karl Marx*, edited by Max Eastman. New York: Modern Library.
- . (1894) 1959b. "Capital, Volume 3, Chapter 9." Pp. 54-59 in *Capital, the Communist Manifesto, and Other Writings by Karl Marx*, edited by Max Eastman. New York: Modern Library.
- Medoff, James J., and Katherine Abraham. 1981. "Are Those Paid More Really More Productive?" *Journal of Human Resources* 16:186-216.
- Mitchell, Daniel J. B. 1985. "Shifting Norms in Wage Determination." *Brookings Papers on Economic Activity*. Washington, D.C.: Brookings Institute.
- Murphy, Raymond. 1988. *Social Closure: The Theory of Monopolization and Exclusion*. Oxford: Clarendon Press.
- Myles, John, Garnett Picot, and Tedd Wannell. 1988. "The Changing Wage Distributions of Jobs, 1981-86." *Canadian Economic Observer* 4:4.2-4.12.
- North, Douglass, and Robert P. Thomas. 1973. *The Rise of Western Capitalism*. New York: Cambridge University Press.
- Olson, Mancur. 1965. *The Logic of Collective Action*. Cambridge, Mass.: Harvard University Press.
- Ossowski, Stanislaw. 1963. *Class Structure in the Social Consciousness*. New York: Free Press.
- Parkin, Frank. 1979. *Marxism and Class Theory*. New York: Columbia University.
- Poulantzas, Nicos. 1975. *Classes in Contemporary Capitalism*. London: New Left Books.
- Ricardo, David. (1821) 1951. *On the Principles of Political Economy and Taxation*. Vol. 1 of *The Works and Correspondence of David Ricardo*, edited by Piero Sraffa. Cambridge: Cambridge University Press.
- Roemer, John. 1982. *A General Theory of Exploitation and Class*. Cambridge, Mass.: Harvard University Press.
- . 1986. "Should Marxists Be Interested in Exploitation?" Pp. 260-82 in *Analytical Marxism*, edited by John E. Roemer. New York: Cambridge University Press.
- Ryan, William. 1971. *Blaming the Victim*. New York: Orbach & Chambers.
- Ryscavage, Paul, and Peter Henle. 1990. "Earnings Inequality Accelerates in the 1980's." *Monthly Labor Review* 113, no. 12: 3-16.
- Schneider, Louis, and Sverre Lysgaard. 1953. "The Deferred Gratification Pattern." *American Sociological Review* 18, no. 2 (April): 142-49.
- Simmel, Georg. 1908. *Soziologie*. Leipzig: Duncker & C. Humblot.
- Simon, Herbert. 1957. "The Employment Relation." In *In Models of Man*. New York: Wiley.
- Solow, Robert M. 1979. "Another Possible Source of Wage Stickiness." *Journal of Macroeconomics* 1 (2): 79-82.
- Sørensen, Aage B. 1983. "Processes of Allocation to Open and Closed Positions in Social Structure." *Zeitschrift für Soziologie* 12:203-24.
- . 1991. "On the Usefulness of Class Analysis in Research on Social Mobility and Socioeconomic Inequality." *Acta Sociologica* 34 (2): 71-88.
- . 1996. "The Structural Basis of Social Inequality." *American Journal of Sociology* 101, no. 5 (March): 1333-65.

- . 1998. "On Kings, Pietism and Rent-Seeking in Scandinavian Welfare States." *Acta Sociologica* 41, no. 4 (August 1998): 363–376.
- Tullock, Gordon. 1980. "The Transitional Gains Trap." Pp. 211–21 in *Toward a Theory of the Rent Seeking Society*, edited by James S. Buchanan, Robert D. Tollison, and Gordon Tullock. College Station: Texas A&M University Press.
- . 1989. *The Economics of Special Privilege and Rent Seeking*. Boston: Kluwer Academic Publishers.
- Warner, W. Lloyd, Marchia Meeker, and Kenneth Bells. 1949. *Social Class in America*. New York: Science Research Associates.
- Weber, Max. (1920) 1978. *Economy and Society*. Berkeley: University of California Press.
- . 1946. "Class, Status and Power." Pp. 180–95 in *From Max Weber: Essays in Sociology*, edited by Hans H. Gerth and C. Wright Mills. New York: Oxford University Press.
- Wright, Erik O. 1979. *Class Structure and Income Determination*. New York: Academic Press.
- . 1982. "The Status of the Political in the Concept of Class Structure." *Politics and Society* 11 (3): 321–52.
- . 1985. *Classes*. London: New Left Books.
- . 1997. *Class Counts*. Cambridge: Cambridge University Press.
- Wolff, Edward N. 1995. *Top Heavy: A Study of Increasing Inequality of Wealth in America*. Washington, D.C.: Brookings Institution.

Class, Exploitation, and Economic Rents: Reflections on Sørensen's "Sounder Basis"¹

Erik Olin Wright

University of Wisconsin, Madison

Aage Sørensen, in "Toward a Sounder Basis for Class Analysis," argues that Marxists are correct in placing exploitation at the center of class analysis since an exploitation-centered concept of class has a much greater potential for explaining the structural foundations of social conflicts over inequality than does its principle rival, the material "life conditions" conception of class.² But he also believes that existing concepts of exploitation are seriously compromised due to an absence of rigorous theoretical foundations. To solve this problem he proposes rehabilitating the concept of exploitation by closely identifying it with the economic concept of rent. This, he believes, retains the fundamental sociological meaning of exploitation while giving the concept much more theoretical precision and analytical power.

I share with Sørensen the commitment to reconstructing an exploitation-centered concept of class (Wright 1979, 1985, 1989, 1997). And, like Sørensen, I believe that a rigorous concept of exploitation can be elaborated without the use of the labor theory of value. I have also argued that there is a close link between the concept of economic rent and various forms of exploitation. I disagree, however, that exploitation can be fruitfully defined simply in terms of rent-generating processes or that a class

¹ Direct correspondence to Erik Olin Wright, Department of Sociology, University of Wisconsin, 1180 Observatory Drive, Madison, Wisconsin 53706-1393.

² Sørensen proposes a threefold classification of class concepts: purely *nominal* class concepts define classes in terms of arbitrary demarcations in systems of stratification; *life conditions* concepts of class "make claims about the empirical existence of observable groupings with identifiable boundaries" sharing common material conditions of existence (Sørensen, p. 1526); *exploitation* concepts of class define classes as conflict groups with inherently antagonistic interests. To this typology I would make two further subdistinctions: within the life conditions concept of class I would distinguish between class concepts that are built around common material conditions of life as such and those that are built around common *opportunity for achieving* material conditions of life (or what is sometimes called "life chances"); within the exploitation concept of class I would distinguish between class concepts that are built around antagonistic interests over material advantages as such, and concepts that emphasize the *interactive interdependency* of classes. This article is primarily about this distinction within the family of exploitation-centered class concepts.

analysis built on such foundations will be satisfactory. The objective of this article is to explain why I feel rent alone does not provide a "sounder basis" for class analysis.

In the next section I will briefly summarize the central ideas of Sørensen's proposal. This will be followed by an explication of an alternative conceptualization of exploitation which sees exploitation as not simply rent-generated advantages, but advantages that involve the appropriation of labor effort of the exploited by exploiters. The article will conclude with a discussion of the complex relationship between rent and exploitation.

SØRENSEN'S MODEL OF RENT-BASED EXPLOITATION

Sørensen begins by endorsing what might be called the root meaning of exploitation in Marx: "Exploitation, for Marx and in this discussion, means that there is a causal connection between the advantage and disadvantage of two classes. This causal connection creates latent antagonistic interests that, when acted upon as a result of the development of class consciousness, create class conflict" (Sørensen, p. 1524). The pivot of this definition is the idea of *antagonistic interests*: "Interests may be said to be antagonistic when the gain of one actor, or a set of actors, excludes others from gaining the same advantage" (p. 1524). An exploitation-centered concept of class, therefore, sees class relations as structured by processes of exploitation that causally generate antagonistic interests.

The central problem, then, is figuring out what properties of social relations in fact generate such antagonistic interests. Sørensen argues, correctly I believe, that the traditional Marxist strategy of basing such an account in the labor theory of value is unsatisfactory. He proposes a simple and straightforward alternative by identifying exploitation with economic rents. Owning assets of various sorts gives people a stream of income—call this returns on owning the asset—when those assets are deployed in production or exchanged in a market. The "value" of the asset to an individual is defined by the total returns one obtains from that asset during the period in which one owns it. We can then define a special counterfactual: the returns to the asset under conditions of "perfect competition." Any return to the asset above this counterfactual is a rent: "The difference between the actual rental price and the competitive price is what is called an *economic rent*. . . . Rents are payments to assets that exceed the competitive price or the price sufficient to cover costs and therefore exceeding what is sufficient to bring about the employment of the asset" (Sørensen, p. 1536).

Perfect competition is a quite demanding condition. It implies *perfect information* and a complete *absence of any power relations* between actors

within a market. Economists are used to including the power condition in discussions of competitive markets. This is where the contrast between competitive and monopolistic markets comes from. Much less attention is generally paid to the information conditions. If actors in a system of exchange and production have incomplete information, the contracts are not costlessly enforceable (since resources must be devoted to monitoring compliance with contracts). In general in such situations the empirical prices in exchange relations will deviate from the prices that would pertain under conditions of perfect information, thus generating rents associated with transaction costs.

With this standard economic definition of rents in hand, Sørensen then proposes to define exploitation in terms of rents: "I propose . . . to restrict exploitation to inequality generated by ownership or possession of *rent-producing assets*. Rent-producing assets or resources create inequalities where the advantage to the owner is obtained at the expense of nonowners. These nonowners would be better off if the rent-producing asset was redistributed or eliminated" (p. 1532; emphasis in original). *Exploitation class* is thus defined as "structural locations that provide rights to rent-producing assets" (p. 1525).

This definition of exploitation class produces some startling conclusions that run quite counter to the conventional intuitions of most class analysts:

1. Capitalist property relations by themselves do not generate classes. "In perfectly competitive markets, with no transaction costs, there are no permanent advantages, or above-market returns, to be obtained at the expense of somebody else. Thus class location is irrelevant" (pp. 1527–28). Strictly speaking, within Sørensen's framework the claim here should be even stronger: it is not simply that class location would be irrelevant; there would be no class locations at all. A capitalist market economy with perfect competition would be a classless society.
2. When labor unions negotiate "solidarity wages" in which wage differentials are reduced by raising the wages of unskilled workers, unskilled workers become an exploiting class: "The main effect of unions is to reduce wage inequality. Unions are especially effective at decreasing the wage spread between more and less productive workers. Unions may create substantial rents to low-skilled or otherwise less productive workers" (p. 1550).
3. The existence of a high minimum wage increases exploitation in a society and renders workers at or near the minimum wage an exploiting class.
4. A strong welfare state also increases exploitation; welfare recipients are an exploiting class.

5. While some of the increase in inequality in the last two decades in many capitalist societies may reflect a redistribution of rents from one category of actors to another (particularly when capitalists are able to capture a higher proportion of what Sørensen calls "composite rents"), mostly this increasing inequality reflects a reduction in rents: "The increase in inequality is very much driven by an increase in wages and earnings of the highest paid workers and stagnation or decline for others. The stagnation and decline follow from the rent destruction" (p. 1552). This implies less exploitation in the system under neoliberal, deregulated labor markets and thus a move toward a classless society.

Sørensen recognizes that these conclusions are deeply counterintuitive. He responds by arguing that in capitalist societies reducing exploitation may in fact be a bad thing for many ordinary people. "Nothing," he writes, "guarantees that efficient labor markets create good lives. Rents are required in modern society to provide decent standards of living for the poorest part of the population. These rents are provided from the state in the form of income support and other welfare goods" (p. 1553). Ironically, then, the elimination of class and exploitation increases human misery, and thus a humane capitalist society is one that fosters certain kinds of antagonistic class interests, particularly by strengthening state-sanctioned forms of exploitation.

There is nothing intrinsically wrong with bold, provocative, counterintuitive claims. Indeed, the hallmark of the best sociology is discovering properties of social relations that go against conventional wisdom and thus counteract the definition of sociology as the painful elaboration of the obvious. Nevertheless, when such striking counterintuitive claims are made they may suggest that there are problems and missing elements in a theoretical proposal. This, I will argue, is the case in the simple equation of exploitation and rents.

AN ALTERNATIVE DEFINITION OF EXPLOITATION

The definition of exploitation I will elaborate shares much with that proposed by Sørensen. Like Sørensen, I argue that exploitation generates antagonistic interests in which the material welfare of exploiters is causally dependent upon harms to the material interests of the exploited. I also believe that this causal dependence is rooted in the ways in which productive assets of various sorts are owned and controlled. And, like Sørensen, I argue that defining class in terms of exploitation rather than simply material conditions of life provides the richest conceptual foundations for linking an account of material inequality with an account of social conflict.

We differ in two respects: first, I do not think that rents in and of themselves provide a full account of the explanatory mechanisms of exploitation, and second, I think capitalism generates antagonistic class interests even under the imaginary conditions of perfect competition. The first of these points involves examining the ways in which exploitation involves the appropriation of "labor effort" rather than simply "advantage"; the second involves showing how capitalist property relations generate antagonisms even under perfectly competitive markets.

Exploitation and the Appropriation of Labor Effort

Exploitation, as I will define the concept, exists when three criteria are satisfied (see Wright [1997, pp. 9–19] for details):

1. *The inverse interdependent welfare principle.*—The material welfare of exploiters causally depends upon the reductions of material welfare of the exploited.³

2. *The exclusion principle.*—This inverse interdependence of the welfare of exploiters and the exploited depends upon the exclusion of the exploited from access to certain productive resources.

3. *The appropriation principle.*—Exclusion generates material advantage to exploiters because it enables them to appropriate the labor effort of the exploited.

Exploitation is thus a diagnosis of the process through which certain inequalities in incomes are generated by inequalities in rights and powers over productive resources: the inequalities occur, in part at least, through the ways in which exploiters, by virtue of their exclusionary rights and powers over resources, are able to appropriate labor effort of the exploited.⁴ If the first two of these principles are present, but not the third,

³ It is often noted that in a market economy both parties to an exchange gain relative to their condition before making the exchange. This applies to ordinary market exchanges of commodities and also to the employment exchange: both workers and capitalists gain when an exchange of labor power for a wage occurs. Such mutual gains from trade can occur and it can still be the case that the magnitude of the gains from trade of one party is at the expense of another party. As it has often been noted (in a quote attributed to the British economist Joan Robinson), "The one thing worse for workers than being exploited in capitalism is not being exploited in capitalism." This general point applies to Sørensen's conception of exploitation-as-rents as well to the conception being proposed here: in situations in which capitalists obtain monopoly rents in the market it is still the case that there are mutual gains from trade by the people who purchase the products at monopolistic prices.

⁴ "Appropriation of labor effort" can take many forms. Typically this involves appropriating the products of that labor effort, but it may involve a direct appropriation of labor services. The claim that labor effort is appropriated does not depend upon the thesis of the labor theory of value that the value of the products appropriated by capitalists is determined by the amount of labor those products embody. All that is

what might be termed *nonexploitative economic oppression* may exist, but not exploitation. The crucial difference is that in nonexploitative economic oppression, the advantaged social category *does not itself need* the excluded category. While the welfare of the advantaged does depend upon the exclusion principle, there is no ongoing interdependence of their activities with those of the disadvantaged. In the case of exploitation, the exploiters actively need the exploited: exploiters depend upon the effort of the exploited for their own welfare.

Sørensen explicitly rejects this third criterion and thus rejects the proposed distinction between exploitative and nonexploitative oppression. He writes:

Wright (1997) proposes a related definition of exploitation though it is not formulated in terms of the concept of rent. In addition to the causal link between advantages and disadvantages of classes, Wright requires that the advantaged class depend on the fruits of labor of the disadvantaged class for exploitation to exist. Thus when the European settlers displaced Native Americans they did not exploit by obtaining an advantage at the expense of native Americans; they engaged in "nonexploitative economic oppression" (Wright 1997, p. 11). The European settlers clearly created antagonistic interests that brought about conflict, so it is not clear what is added by the requirement of transfer of the fruits of labor power.

One way of seeing "what is added by the requirement of transfer of the fruits of labor" is to contrast historical situations in which exploitation occurs with those characterized by nonexploitative oppression. Consider the difference in the treatment of indigenous peoples in North America and South Africa by European settlers. In both places the first two criteria above are satisfied: in both there is a causal relationship between the material advantages of settlers and the material disadvantages of indigenous people; and in both this causal relation is rooted in processes by which indigenous people were excluded from a crucial productive resource, land. In South Africa, however, the third principle was also present: the settler population appropriated the fruits of labor of the indigenous population, first as agricultural labor working the land and later as mine workers, whereas in North America the labor effort of indigenous people was generally not appropriated.

Does this matter? It does not matter, perhaps, if all we are concerned with is the sheer presence or absence of "antagonistic interests," for in both instances there surely was deep antagonism. But the dynamics of the antagonism are fundamentally different in the two cases: in North

claimed is that when capitalists appropriate products they appropriate the laboring effort of the people who make those products.

America, because the settler population did not need Native Americans, they could adopt a strategy of genocide in response to the conflicts generated by the exclusion of indigenous people from the land. There is a morally abhorrent folk expression in American culture that reflects the specificity of this antagonism: "The only good Indian is a dead Indian." No comparable expression exists for workers, slaves, or other exploited classes. One might say "the only good slave is a docile slave" or "the only good worker is an obedient worker," but it would make no sense to say "the only good worker is a dead worker" or "the only good slave is a dead slave." Why? Because the prosperity of slaveowners and capitalists depend upon the expenditure of effort of those whom they exploit. Sørensen's definition of exploitation does not distinguish between what I call exploitative and nonexploitative oppression and thus does not capture this strong sense in which *exploiters depend upon and need the exploited*.

This deep interdependence makes exploitation a particularly explosive form of social relation for two reasons: first, exploitation constitutes a social relation that simultaneously pits the interests of one group against another and that requires their ongoing interactions; second, it confers upon the exploited group a real form of power with which to challenge the interests of exploiters. This is an important point. Exploitation depends upon the appropriation of labor effort. Because human beings are conscious agents, not robots, they always retain significant levels of real control over their expenditure of effort. The extraction of effort within exploitative relations is thus always to a greater or lesser extent problematic and precarious, requiring active institutional devices for its reproduction.⁵ Such devices can become quite costly to exploiters in the form of the costs of supervision, surveillance, sanctions, and so forth. The ability to impose such costs constitutes a form of power among the exploited.

My first conclusion, then, is that a concept of exploitation based solely on the notion of rent misses the ways in which exploiters are not merely advantaged because of the disadvantages of the exploited but are dependent upon the exploited. This dependency is a central feature of class relations.

⁵ The claim that, because of the antagonistic interdependency of material interests generated by exploitation, class relations *require* active institutional devices for their reproduction is at the core of the Marxist tradition of class analysis. This claim provides the basis for the attempt at building an endogenous theory of ideology and the state, either in the strong form of the base/superstructure theory of classical historical materialism or weaker forms of contemporary neo-Marxism. The theory predicts that if class relations are to be reproduced stably, there will be a tendency for forms of ideology to emerge that mask that exploitation and for forms of the state to emerge that obstruct efforts at challenging those class relations.

Class, Exploitation, and "Perfect Competition"

In Sørensen's rent-centered concept of class and exploitation, a capitalist market economy with perfect competition (which also requires perfect information) would be a classless society since all returns to assets would be exactly equal to their costs of production.⁶ There would therefore be no rents and thus no exploitation or class.

Traditional Marxist conceptions of class and exploitation are sharply at odds with this diagnosis. In the framework of the labor theory of value it was easy to demonstrate that workers were exploited even under conditions of perfect competition. What happens when we abandon the labor metric of value?

Let us examine the three criteria for exploitation specified above in a capitalist economy with perfect competition in which there are only two categories of economic actors: capitalists who own the means of production—and thus have the effective power to exclude others from access to those assets—and workers who own only their labor power. We will not call these "classes" yet because we first need to see if the three criteria for exploitation are satisfied. For simplicity, let us assume that capitalists are pure *rentiers*: they invest their capital in production and receive a rate of return on those capital assets, but they do not themselves work (managers are thus simply a type of worker in this simple case). Are the inverse interdependent welfare principle and the exclusion principle satisfied in this case? Is the material welfare of capitalists causally dependent upon the exclusion of workers from access to capital assets? The test here is whether or not it is the case that workers would be better off and capitalists worse off if property rights were redistributed so that workers would no longer be "excluded" from capital.⁷ It seems hard to argue that this is not the case: in the initial condition capitalists have a choice of either consuming their capital or investing it, as well as the choice of whether or not they will work for earnings. Workers only have the latter choice. To be sure, they can borrow capital (and in a world of perfect information

⁶ The concept of "costs of production" in this context includes such things as the "costs of deferring present consumption for greater future consumption." Thus, in a perfectly competitive market with perfect information, the interest rate on loans and the profit rate on investments are both simply the necessary returns on the relevant assets needed to compensate exactly the owners of those assets for the "costs" of foregone consumption. (There are further complications introduced by issues of risk, but the basic idea is the same).

⁷ This is the test John Roemer (1983) proposes for assessing the existence of what he terms *capitalist exploitation*: would workers be better off and would capitalists be worse off if workers left the "game" of capitalism with their per capita share of capital. His answer is that in general they would be better off, and this demonstrates that within capitalism itself capitalists' welfare occurs at the expense of workers.

they would not need collateral to do so since there would be no transaction costs, no monitoring costs, no possibility of opportunism), but still workers would be better off owning capital outright than having to borrow it.

In his analysis of rent-based exploitation, Sørensen argues that where rent-producing assets exist “nonowners would be better off if the rent-producing asset was redistributed or eliminated” (p. 1532). What I have just noted is that where capital assets are privately owned and unequally distributed—in particular, where one group of agents has no capital assets and another group has sufficient capital assets to not need to work—then, even if those assets do not generate rents it is still true that “nonowners would be better off if the income-producing asset was redistributed.”⁸

What about the third criterion for exploitation, the appropriation principle? Does it make sense to say that in a system of perfect competition with perfect information and equilibrium prices, *rentier* capitalists appropriate the “labor effort” of workers. As Sørensen correctly points out, when such an imaginary system is in equilibrium, workers who expended more effort would be paid more, workers who shirked would be paid less, and (according to standard marginalist reasoning) the amount they were paid for their effort level would exactly reflect the price of the product they produced with that effort. This is the kind of reasoning that has always lead neoclassical economists to deny the existence of exploitation in capitalism.

Nevertheless, even under these conditions the following is true: (1) the only labor effort performed in the system is by workers; (2) capitalists appropriate the product and thus appropriate the “fruits of labor effort” of workers;⁹ (3) for any given wage level capitalists have an interest in getting workers to expend more labor effort than workers would spontaneously want to expend; (4) if workers owned their own means of produc-

⁸ It should be noted that if we have a competitive market in all respects except for the perfect information condition, then the ownership of capital will also generate rents in Sørensen’s sense. That is, as Stiglitz and Weiss (1981) and Bowles and Gintis (1990) have shown, under conditions of imperfect information, interest rates in credit markets will be below the “market clearing rate” (implying that credit is rationed rather than allocated strictly on the basis of price). This implies that profits generated through the use of that credit will contain a rent component, functionally equivalent to the employment rent in nonclearing labor markets. Capitalists who deploy their own capital in production receive this rent as well. Imperfect information is the universal condition in credit markets and thus profits will always have a rent component. Even in Sørensen’s framework, therefore, the sheer ownership of capital should generate rent-based antagonisms of interests.

⁹ This claim does not depend upon the strong claim that the entire value of what capitalists appropriate is a function of the amount of labor embodied in that production. G. A. Cohen (1988) provides a careful defense of the view that the appropriation of the fruits of labor effort as such constitutes exploitation.

tion, capitalists would find it more difficult to get workers to work as hard for a given level of wages. In a purely competitive capitalist economy, therefore, there are still antagonistic interests over the expenditure and appropriation of labor effort.

CLASS, EXPLOITATION, AND RENTS

If one accepts the arguments above, then the simple equation of exploitation with economic rents is an unsatisfactory basis for class analysis. This does not mean, however, that the concept of economic rent is irrelevant to class analysis, but simply that it bears a more complex relationship to the problem of exploitation. Let us examine two specific examples: employment rents derived from transaction costs in the employment contract, and solidarity rents generated by union power.

In Sørensen's analysis, the employment rents workers receive because of transaction costs in the employment relation count as a form of exploitation rendering such workers an exploiting class. In the analysis proposed here, employment rents in general constitute one of the ways in which workers are able to mitigate their own exploitation.

Here is the basic argument: Where imperfect information exists—which is the usual condition of labor contracts—capitalists are generally prepared to pay employment rents to workers in order to extract adequate labor effort from them. The mechanism in play here has been carefully elaborated by Bowles and Gintis (1990) following earlier work on efficiency wages by Akerloff and Yellen (1986) and others: because of imperfect information in the labor market and labor process, capitalists are forced to spend resources on enforcing the labor contract (through supervision, monitoring, etc.) in order to detect shirking. Catching workers shirking is only useful for capitalists if workers care about being punished, especially about losing their jobs. The salience of the threat of being fired for shirking increases as the cost of job loss to workers increase. Paying workers an employment rent—a wage significantly above their reservation wage—increases the cost of job loss and thus the potency of employer threats. The costs to employers to extract labor effort thus consist of two components: the costs of catching workers shirking (monitoring costs) and the costs of making job loss hurt (employment rents).¹⁰ The “employment

¹⁰ As Bowles and Gintis (1990) stress in their analysis of “contested exchange,” the relationship between these two costs can be viewed as a “labor extraction function” within the production process. Employers can increase labor extraction by allocating more resources to monitoring (thus increasing the probability of detecting shirking) or by increasing employment rents. Employers thus, in general, face a strategic trade-off between these two costs.

rent" is thus a wage premium workers are able to get because of their ability to resist capitalist attempts at extracting labor effort. In conditions of perfect (and thus costlessly acquired) information, the capitalist capacity to appropriate effort is enhanced since workers lose this ability to resist. Rather than seeing employment rents as a form of exploitation by workers it is thus more appropriate to see them as the outcome of resistance to exploitation by workers.

As a second example of the relationship of rents and exploitation to class, consider the role of unions in reducing wage differentials among nonmanagerial employees. In Sørensen's provocative characterization of this phenomenon, employed low-skilled workers—the principal beneficiaries of the solidarity wage—are an exploiting class. In terms of the labor-effort appropriation formulation of the concept of exploitation one would want to ask, Who is exploited by these low-skilled workers?¹¹ There are three principle candidates: unemployed low-skilled workers, skilled workers, and capitalists.

An argument can certainly be made that unemployed workers are potentially harmed by solidarity wages. By raising the wages of low-paid workers, employers are likely to hire fewer low-skilled workers than they would do in unregulated labor markets. Solidarity wages—like minimum wages, job security protections, and other such institutional arrangements—create labor market rigidities that create advantages to insiders at the expense of outsiders. But does this warrant the claim that employed low-skilled workers with solidarity wages "exploit" the unemployed? In terms of the proposed definition of exploitation adopted here, they do not. The key question is this: If the unemployed simply disappeared—if they migrated, for example—would the material welfare of the employed low-skilled worker go down or up? If anything, the welfare of employed low-skilled workers would go up if the "reserve army of the unemployed" were to decrease. Unemployed workers in this situation may be subjected to a form of nonexploitative oppression by being denied access to jobs, but they are not exploited.

What about skilled workers? It is certainly the case that in conditions of solidarity wages, the wages of skilled workers are, at least statically, lower than they would be in the absence of union-generated reductions of wage differentials. There are two reasons why it still does not make sense to say that the unskilled exploit the skilled in this context. First, one of the reasons for the solidarity wage is the belief that it enhances overall

¹¹ It should be noted that throughout Sørensen's article there is very little explicit discussion of *who* is being exploited. The general idea seems to be that for any given category of agents who receive rents of one sort or another, the complement of the category is exploited.

class solidarity and thus shifts the balance of power between organized workers and capitalists in favor of workers. In the long run this advances the material interests of both skilled and unskilled workers. Second, if in the absence of the solidarity wage skilled workers would themselves be recipients of union-generated rents, then what the solidarity wage really does is to simply reduce the rents acquired by skilled workers and then redistribute some of these to unskilled workers. Even if one regards rents per se as a form of exploitation, this would not constitute a form of exploitation *of* skilled workers *by* the unskilled, but of the economic agents who pay the rents to the skilled workers in the first place.

If solidarity wages are viewed simply as a redistribution of rents from the skilled to the unskilled, then this suggests that perhaps the unskilled are exploiting capitalists since, after all, capitalists are paying the wages. But capitalists—under the definition of exploitation being used here—are themselves exploiters of workers by virtue of appropriating the fruits of labor of workers. The rent component of the solidarity wage—like the more general wage premium of employment rents—should therefore in general be thought of as a mitigation of capitalist exploitation rather than a form of exploitation in its own right.

Both of these examples show that once the appropriation of labor effort is added as a criterion for the concept of exploitation, the relationship between class, exploitation, and rents becomes much more complex. In some cases rents might still be directly a form of exploitation. The rent that a landowner charges a tenant farmer constitutes a direct appropriation of the labor effort of the farmer, for example. In other cases, rent-acquisition is better thought of as a way of mitigating exploitation. It is for this reason that in general I have argued that employees who are the recipients of various forms of rents within their earnings should be regarded as occupying “privileged appropriation locations within exploitation relations” (Wright 1997, p. 22).¹² The concept of economic rent therefore can play a useful role in the theory of class and exploitation by clarifying the range of mechanisms by which exploitation is enhanced or counteracted, but not by reducing the concept of exploitation simply to advantages obtained by asset owners under conditions of imperfect competition and imperfect information.

¹² This formulation represents a change from my earlier work on class and exploitation (Wright 1985) in which I regarded both skill rents and loyalty rents (rents appropriated especially by managers because of their control over organizational apparatuses) as distinctive forms of exploitation. For a discussion of the reasoning behind this shift, see Wright (1989, pp. 331–40).

REFERENCES

- Akerlof, George, and Janet Yellen. 1986. *Efficiency Wage Models of the Labor Market*. New York: Cambridge University Press.
- Bowles, Samuel, and Herbert Gintis. 1990. "Contested Exchange." *Politics and Society* 18 (2): 165–222.
- Cohen, G. A. 1988. "The Labour Theory of Value and the Concept of Exploitation." Pp. 209–38 in *History, Labour and Freedom*, edited by G. A. Cohen. Oxford: Clarendon Press.
- Stiglitz, Joseph E., and Andrew Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review* 71:393–410.
- Wright, Erik Olin. 1978. *Class, Crisis and the State*. London: New Left Books.
- . 1985. *Classes*. London: Verso
- . 1989. *The Debate on Classes*. London: Verso
- . 1997. *Class Counts: Comparative Studies in Class Analysis*. Cambridge: Cambridge University Press.

Rent, Class Conflict, and Class Structure: A Commentary on Sørensen¹

John H. Goldthorpe
Nuffield College, Oxford University

I much appreciate Aage Sørensen's article, "Toward a Sounder Basis for Class Analysis." I am entirely in sympathy with his argument that, following on the effective collapse of the Marxist tradition, class analysis needs to be informed by better theory. I also agree that attempts to develop theory on supposedly "Weberian" or "neo-Weberian" lines have not been impressive.² More specifically, I find helpful Sørensen's attempt to "bring some order" into class concepts by differentiating them according to the "level of theoretical ambition" with which they are associated (pp. 1525–26). This attempt does in fact provide me with the starting point for my commentary on his article.

Sørensen distinguishes three main types of class concept, hierarchically ordered, that I would myself, with only minor glossing of his own formulations, characterize as follows:

1. Class concepts that provide a purely nominal categorization of populations, according to a particular "dimension," or possibly a combination of dimensions, of stratification and that can be used to display class inequalities in life conditions and life chances or class differences in attitudes, beliefs, values, and patterns of action
2. Class concepts that go beyond point 1 above in aiming to delineate sets of class positions among which individuals or families within a population are distributed and may thus come to form collectivities with a greater or lesser degree of demographic, subcultural, and social identity
3. Class concepts that go beyond 1 and 2 in aiming to specify actual

¹ I am indebted to Robert Erikson for helpful conversation and for comments on an earlier draft. Direct correspondence to John Goldthorpe, Nuffield College, Oxford University, Oxford OX1 1NF, England.

² My own approach to class analysis has often been characterised as Weberian, but I would not myself regard this labeling as being very informative. I would in fact share Sørensen's view (p. 1527, n.3) that the importance given to Weber in the literature on class analysis is "a bit curious."

collectivities within populations whose members have the motivation to engage in conflict with members of other comparable collectivities in consequence of their interests being structurally opposed; that is, opposed by virtue of the class positions that they hold

Thus, those sociologists deploying type 2 concepts tend to be concerned not just with revealing class inequalities or class differences but further with explaining the conditions under which varying degrees of class formation occur and, in turn, with giving accounts of what Sørensen nicely calls (p. 1526) "the geography of social structure." And those deploying type 3 concepts seek yet further to explain not just how inequalities and differences arise between classes but also antagonistic interests that create the potential for conflict.

In his article, Sørensen clearly aims to provide "a sounder basis for class analysis" by developing type 3 concepts and pursuing the theoretical ambitions that they are devised to serve. But, if I understand him correctly, he would—consistent with the hierarchical nature of his typology—*also* wish the theory that he outlines to be one that can contribute to our understanding of the geography of social structure and to the analysis of the nature and extent of class inequalities and differences. At all events, I would myself want to argue that any well-constructed theory of social class should in fact have the potential to operate effectively at all three of the levels of theoretical ambition that Sørensen specifies.

My own work in this connection began with type 2 class concepts and related theoretical concerns that arose out of a dissatisfaction with type 1 concepts (in particular, with that of "socioeconomic status," as used in much American sociology). Together with a number of colleagues, I elaborated a "class schema," for use in empirical research, which derived from the idea that in modern societies class positions are best understood as being defined by differences in employment relations (see esp. Erikson and Goldthorpe 1992; Goldthorpe 1997a).³ However, as Sørensen notes (p. 1527, n.2), I have more recently—encouraged by results from studies of the criterion as well as the construct validity of the class schema (Evans 1992, 1996; Evans and Mills 1998)⁴—moved on to try to explain just why a

³ In consequence of its rather complex genesis, the schema is known by several different names. In a British context, it is usually referred to as the Goldthorpe schema; in an international context, as the EGP (Erikson-Goldthorpe-Portocarero) schema, the Erikson-Goldthorpe schema, or the CASMIN schema, after its use in the Comparative Analysis of Social Mobility in Industrial Nations project.

⁴ I refer here to studies that show that not only does the class schema perform well in displaying theoretically expected class inequalities and differences but further that, as operationalized through information on occupation and employment status, it does

class structure comprising positions differentiated in terms of employment relations *should be* an apparently abiding feature of modern societies and give rise to persisting class inequalities and differences and further, subject to certain conditions and limits, to various expressions of class conflict (Goldthorpe 2000, chap. 10).⁵

In what follows, I set out two main reservations that I have regarding the adequacy of Sørensen's proposals to serve as the basis of class theory of a "comprehensive" kind in the sense I have indicated, and I seek to develop my arguments by reference to both similarities and differences between Sørensen's approach to class analysis and my own. Finally, I note how our two approaches lead to contrasting expectations (all else being equal) regarding the future of the class structures of modern capitalist societies, so that the possibility exists of some eventual empirical assessment of their respective merits.

My first reservation concerns the conceptual viability, from the standpoint of a comprehensive class theory, of taking what Sørensen calls the "rent distribution" within a society (as distinct from mere class inequalities in "total wealth") as being the source of opposing class interests and in turn of the formation of antagonistic classes—exploitation classes—that engage in collective action directed toward the creation and preservation or the destruction of rents.⁶ I would not wish to deny that action of the kind in question is widespread in modern societies, and I would indeed agree with Sørensen that sociologists should pay more attention to it than they do. However, what I would still query is whether it is helpful *either*

in fact capture in a generally satisfactory way the differences in employment relations that it is conceptually supposed to capture.

⁵ This chapter in *On Sociology* is a revised version of Goldthorpe (1997b) as cited by Sørensen (n. 2). I would add to his note that as well as drawing on transaction cost economics I draw to at least an equal extent on recent work in organizational and personnel economics. My efforts in developing the theoretical basis of the class schema were stimulated by the fact that, following reports of a committee of the British Economic and Social Research Council, set up on behalf of the Office of National Statistics (Rose and O'Reilly 1997, 1998), the schema was accepted as the basis for a new official social classification for the United Kingdom, which has now replaced the Registrar General's Social Classes, in operation since 1911.

⁶ I am puzzled by Sørensen's use of the concept of exploitation—a word that I would myself gladly see disappear from the sociological lexicon—since it is not clear that it plays any vital role in his arguments. Its function in Marxist thought was to allow a fusion of normative and positive claims in a way that I would find unacceptable, as also, I believe, would Sørensen. For example, it is evident that he would regard the elimination of rents, and thus of exploitation, from labor markets as tending often to have highly *negative* human consequences. If invoking exploitation is no more than a way of flagging the presence of structurally opposed class interests that lead to zero-sum conflicts, then its use is innocuous but scarcely necessary.

to regard *all* class conflict as being conflict over rents *or* to regard all conflict over rents as being conflict *between classes*.

As regards the first proposition, my difficulties arise from the (very broad) definition of rent that Sørensen adopts: rent, it would appear, is any return to an asset that exceeds the return that would be gained in an economy operating under conditions of perfect competition—as these would be defined in a standard economics textbook. The perfectly competitive economy is Sørensen's baseline for determining whether rents exist. There are, then, two important implications of Sørensen's position, and ones which—I think—he is ready to accept. First, since rents mark, as it were, deviations from the functioning of a perfectly competitive economy, they must lead to a shortfall from the standard of efficiency that such an economy sets. And second, since all class conflict is conflict over rent, it must be the case that in a perfectly competitive economy no class conflict could arise.

However, the perfectly competitive economy is an abstraction. Outside of economics textbooks, economies do not exist and function *in vacuo* but as "embedded" in complex institutional contexts. And in modern capitalist societies, I would then argue, it is in fact over the particular form that such supporting institutions should take that class conflict very largely occurs: most prominently, perhaps, in regard to social welfare rights and provisions and the manner of their financing, as these bear upon the degree of class inequalities of both condition and opportunity.

Sørensen would, of course, wish to treat all such conflicts over welfare rights and such as being conflicts over rents. But an important issue then is that of how far welfare and other institutions, and the rents (in Sørensen's sense) associated with them, *can* be understood simply as representing so many deviations from the principles of a perfectly competitive economy and therefore as sources of economic inefficiency—or have, rather, to be seen as significantly contributing to the viability of any real world economy that seeks to approximate the competitive ideal. Most social democratic interpretations of the "institutional" (as opposed to the "residual" or "marginal") welfare state would see this as an achievement of the "democratic translation" of the class struggle that in fact promotes more than it undermines economic efficiency. An advanced welfare state serves not so much as a "leaky bucket" of rent redistribution but more as a vital "irrigation system" that prevents the waste of human capital and the underutilization of productive resources in general (Korpi 1983, 1985; see also Rieger and Leibfried 1998). Unless, then, Sørensen wishes totally to reject such interpretations—and he makes no evident effort in this direction—his restriction of class conflict to conflict over rents would seem somewhat artificial and, in so far as the creation rather than the destruc-

tion of rents can be taken as serving efficiency, risks losing theoretical coherence.⁷

The second proposition—that all conflicts over rent are conflicts between classes—I find yet more problematic in its implications. As Sørensen makes very clear, conflicts over rent, in his sense, are highly heterogeneous. They are such that they can readily arise *among* capitalists and *among* workers, even workers of similar grade, as well as *between* capitalists and workers. Thus, capitalists (or their agents) may engage in conflict over the securing or terminating of monopolies, licenses, or patents that afford rents; likewise, workers may engage in conflict over the rents that can be obtained from organization or from involvement in firm-internal labor markets or from minimum-wage legislation.

Now Sørensen would presumably wish to argue that we should be ready to break with our established ways of thinking and to treat such conflicts as indeed being class conflicts, no less than those that divide capitalists and workers; he is of course fully entitled to propose such a conceptual innovation. However, it is then difficult to see how a theory of class could be built on this basis that would also be adequate to the *other* concerns of class analysis previously noted: that is, in regard to class inequalities and differences and the formation or decomposition of classes as features of the geography of social structure. Rather obviously, the lines of class conflict over rents that would be drawn among capitalists (and their agents) and among workers would not, in the same way as those drawn between capitalists and workers, map onto the main lines of class inequalities (including those in “total wealth”) or of class differences that can be empirically established, more or less clearly, on the basis of any existing type 1 or type 2 concepts. Would it be possible, following Sørensen’s approach to class analysis, to arrive at type 1 or type 2 concepts that would in fact be serviceable in empirical research?

Moreover, the heterogeneity of rents and of the conflicts to which they give rise would serve to inhibit the degree of formation of Sørensen’s “exploitation classes,” as in turn would the degree of shifting and cross-cutting of the lines of conflict that would surely be found. Many individuals would be exploited in one context, but act as exploiters in another. Thus, within modern societies at least, it would seem unlikely that conflict over rents could ever become conflict on a “totalizing,” societal scale. Sørensen

⁷ These considerations lead me to suspect that Sørensen might have been better served by a somewhat more restricted understanding of rent. The history of the concept in economic analysis is indeed one of progressive generalization. But a comment of Schumpeter (1954, p. 679) on an earlier stage of this process is of interest: “There are generalizations that spell additional success for a theory: they enrich and extend, but do not endanger, its original interval of application. But there are others that spell or foreshadow the break-up of the theory.”

does in fact appear to recognize this, at least as regards conflict that has the potential for radical social transformation. For he remarks (p. 1541) that "class conflict should be more prevalent under feudalism than capitalism" and then goes on directly to observe, in support of this view, that "no revolution has occurred in an advanced industrial society."

However, for non-Marxists at least, it would not appear either necessary or desirable to envisage all class conflict as being directed toward revolution. I would myself regard the occurrence of certain forms of class conflict as being entirely compatible with the acceptance by all parties involved of the institutions of liberal democracy; or, in other words, I would recognize the possibility—and, in some historical instances, the actuality—of the "democratic translation" of the class struggle, as previously referred to. At the same time, though, I would still wish to preserve the conventional distinction that Sørensen's notion of "exploitation classes" opposing each other over rents would seem necessarily to blur: that is, between class conflict that through the solidaristic transcendence of special interests can possibly find expression at a societal level and "interest-group" conflict that remains in its nature sectional and localized.⁸

My second reservation about Sørensen's proposed "sounder basis for class analysis" concerns the reluctance he shows—along with most neo-classical economists—to accepting the distinctiveness of the labor market and, more specifically, of the employment contract. For Sørensen, "employment rents" are not essentially different from other rents. They arise insofar as a labor market is in some way "closed" rather than being "open," that is, perfectly competitive. Such closure may result from the restrictive practices of unions or professional associations or from various kinds of governmental intervention; but it may also result directly from employers entering into certain kinds of employment contracts, chiefly ones that reflect the bargaining advantages that workers enjoy via problems of work monitoring or human asset specificity. Employment rents then lead, just like all other rents, to conflict between classes (in Sørensen's sense) that could only be resolved by the destruction of such rents, whatever their origin might be, and by labor markets becoming open in all respects: that is, becoming ones in which there is no restriction on supply, no elements of subsidy and so forth, *and* in which all employment contracts take on a "rent-free" form. This would then, so far as I can see (though Sørensen is not himself explicit on the point), imply straightforward "spot" contracts

⁸ On the basis of this distinction, there has of course been much discussion, especially in regard to the working class, of the way in which interest-group activity, including activity that could clearly be regarded as "rent seeking," inhibits class formation and reduces the possibility of effective strategic action being taken at a societal level, through class organizations.

in which so much labor, treated simply as a commodity, is exchanged for so much money, with renewal of the contract being entirely contingent. All other forms of employment contract, in their propensity to give rise to rents, would be inefficient: that is, they would fall short of the standard set by perfect competition.

In the approach to class analysis that I would myself follow, recognition of the distinctiveness of the employment contract has, in contrast, a crucial role. Because labor cannot be physically separated from the persons who provide it, what is bought and sold through this contract, I would argue, is not a commodity, in the sense of some definite, objective thing, but rather a *social relationship*. Employment contracts are contracts through which employees agree, in return for remuneration in some form, to place themselves under the authority of an employer or of the employer's agents. A consequences of this, I would then hold, is that spot contracts need not be, and often in fact are not, the most efficient form of contract.

The initial lines of demarcation within the class schema that my colleagues and I have developed are those drawn—following the general idea of class positions being defined by employment relations—between employers, self-employed workers, and employees. But the numerically preponderant category of employees is further differentiated according to the kinds of contracts on the basis of which their work is regulated. At one pole are those employees, most characteristically unskilled manual and entirely routine grades of nonmanual worker, whose work is regulated by the basic “labor contract” that in fact comes as close as is possible to a spot contract in the sense previously referred to. At the other pole, however, are those employees, most characteristically professionals, administrators, and managers, whose work is regulated by a contract of a significantly different kind establishing a “service relationship.” While the basic labor contract entails a specific wage-based, “money-for-effort” bargain, the service relationship entails a more diffuse and longer-term exchange in which, in return for service to their employing organization, employees receive “compensation” that comprises important prospective elements relating to salary increments, promotion, employment security, and so on. Modified versions of the basic labor contract and service relationship are also recognized, as are “mixed” forms of contract, and these can likewise be associated with different types of work (Goldthorpe 2000; see chap. 10 for a fuller exposition).

I would then see this variation in the form of employment contracts as providing the main basis for explaining why the class schema in general performs well in displaying theoretically expected inequalities and differences—that is, the class schema can claim good construct validity. Members of the classes it distinguishes, as well as having differing sources and levels of income, also have differing degrees of stability of both income

and employment and differing expectations as to their economic futures that together condition both their life chances and many aspects of their attitudes and patterns of action. Moreover, the class schema serves well to display the rates and patterns of inter- and intragenerational class mobility that are crucial to processes of class formation (or decomposition), at least at a “demographic” level and, further, the theoretical basis of its construction provides at least a starting point for the explanation of these rates and patterns.⁹

However what, for present purposes, is of chief relevance is the way in which my view of the significance of variation in employment contracts and that taken by Sørensen diverge. As noted, Sørensen would apparently regard all employment contracts that fail to approximate spot contracts—chiefly because of employers’ problems over work monitoring and human asset specificity—as being contracts that give rise to rents and in turn to conflicts of a zero-sum kind (see, e.g., p. 1544). In contrast, I would wish to understand such contracts as being ones that are designed and implemented specifically to overcome problems of the kind in question—problems that inevitably occur with many types of work—and, furthermore, that can overcome them in ways that contribute to organizational effectiveness and *increase the total value of the contract*. Thus, while in the case, say, of contracts that set up a service relationship there may well be Marshallian “higgling and bargaining” over their particular terms, the general form of such contracts offers advantages to employer and employees alike. In turn, then, I would again wish to question the association that Sørensen supposes between rent and inefficiency. From the standpoint of organizational effectiveness, it is in fact the basic labor contract that I would see as being the form of regulation of employment with the most limited range of application: that is, to types of work where little more than labor in its elementary physical sense is involved and employers can thus come as close as is actually possible to “commodification.”

This divergence of view, it should be observed, does then lead on to issues that are of more interest and importance than mere disputes over conceptual preferences (with which, I am uncomfortably aware, much of the foregoing may appear to have been concerned). Sørensen’s contribution has the great merit, in contrast to much other writing on class analysis, of moving from nominal to real propositions: that is, from propositions about how we might see the world to propositions about how the world actually is. More specifically, having argued the case for treating rents as

⁹ The linkage of class, as conceptualized in the schema, to educational decision making is further treated in Goldthorpe (1996) and Goldthorpe and Breen (1997); to voting, in Evans (1993), to health in Bartley et al. (1996), and to mobility in Goldthorpe (2000, chap. 11).

the basis of class division and conflict, he spells out what he takes to be the substantive implications of his position in the context of contemporary capitalist society.

In this regard, a central feature of the scenario that Sørensen presents as already unfolding is a drive by employers, exposed to intensifying globalized competition and subject to the tyranny of stock market valuations, to eliminate employment rents. This they seek to do not only through weakening organized labor or limiting governmental intervention in labor markets, but further by an avoidance of the kinds of employment contract in which such rents are implicit. For example, employers aim progressively to reduce problems of work monitoring and human asset specificity—and the bargaining advantages that these problems give to workers—by introducing new technologies, by redesigning jobs, or by subcontracting or otherwise outsourcing work. It is in the interests of capitalist employers, Sørensen contends (p. 1554), “to produce a labor market conforming to the assumptions of neoclassical economics”—that is, a market in which all employment contracts approximate spot contracts—and this is exactly what they are in process of doing. Thus, the eventual outcome will be that employers will stand opposed to a merely “structureless” mass of workers (p. 1555), within which individuals will be differentiated by the returns they receive to their productive capacities but not by the class positions that they hold in virtue of the particular ways in which their employment relations are contractually mediated. At this stage, therefore, class conflict within the labor market will be at an end.

My own approach to class analysis also gives rise to expectations in the areas on which Sørensen focuses, but to ones of a significantly different kind.¹⁰ I would certainly believe that employers, under the kinds of pressure that Sørensen identifies, will attend closely to contractual arrangements and that the regulation of employment on the basis of the service relationship, where this represents no more than “organisational slack,” is very likely to be attenuated. But I would at the same time regard the scope for such action as being subject to clear limits. In modern business organizations, there will always remain—and whatever degree of downsizing or delayering may be achieved—substantial areas of work, notably of a professional, administrative, and managerial character, where the service relationship in fact best serves productive efficiency. Problems of work monitoring, especially those arising out of complex principal-agent relations, and problems of human asset specificity, especially those of giv-

¹⁰ It may, however, be noted that I share with Sørensen a rational action perspective: i.e., like him, I suppose that capitalist employers, and other parties involved, are, to the best of their ability, acting in a rational—and thus intelligible—way in the pursuit of their goals.

ing incentives to both organizations and workers to invest in training and “planned experience,” create the need for contractual arrangements that have a more sophisticated rationale than that of the discrete, short-term money-for-effort bargain. And even if new technology can in certain respects provide the means of reducing these problems, in others it would seem likely only to accentuate them.

Thus, I would see little possibility of the entropic decline of modern capitalist societies into what, in a seminar presentation, Sørensen once memorably described as “neoclassical soup.” To the contrary, I would regard a relatively complex class structure as being an abiding, integral feature of such societies: class will not cease to be a salient feature of their structural geography. In addition to the primary divisions between employers, self-employed workers, and employees, further, and in many respects yet more important, divisions will persist within the body of employees. That is to say, divisions reflecting differing forms of the regulation of their employment, the logic of which derives ultimately from the distinctive nature that the labor market, even under the most advanced manifestations of capitalism, will still possess.

Of these divisions, the most obvious is that between, on the one hand, the working class, or, more precisely, the class of wage workers, manual and nonmanual, and, on the other, what I would label as the “service class” or salariat of professionals, administrators, and managers (although, as previously intimated, forms of employment contract that mix elements of the basic labor contract and the service relationship can be associated with various “intermediate” class positions). These classes will engage in conflict, whether via individual action—such as voting—or via organization, over what Sørensen would treat as issues of rent but what I would view generically as issues of class inequality of both condition and opportunity. For while the class structure itself will not be dissolved, there will remain significant scope for inequalities between members of different classes either to widen or to narrow, in consequence of political as well as economic actions and circumstances. As but one illustration of this point, it may be noted that the increase over recent decades in earnings and income inequality in the United States, to which Sørensen gives great emphasis, has not been experienced in other capitalist countries, such as France (see further Atkinson 2000). In sum, I venture to suppose that the class schema that my colleagues and I have developed, in capturing differences in employment relations and including those that derive from the differentiation of employment contracts, provides a conceptual basis appropriate to all three levels of theoretical ambition that Sørensen recognizes within the project of class analysis *and* that it will continue to do so for as far ahead as it is reasonable to look.

There remains, of course, the ultimate question of whether it is

Sørensen's prognostications or my own that are gaining most support from the emerging empirical evidence. I cannot take this question up in the space here available to me (but see Goldthorpe 2000, chap. 10), and it might in any event be argued that the evidence of a directly relevant kind that is so far available is still too limited to warrant any strong conclusions. However, thanks to Sørensen's forthright article, and to his readiness to have it openly discussed, the issue has been fairly joined and future developments can be awaited with interest.

REFERENCES

- Atkinson, Anthony B. 2000. "Is Rising Income Inequality Inevitable? A Critique of the Transatlantic Consensus." WIDER Annual Lecture, University of Oslo.
- Bartley, Mel, et al. 1996. "Measuring Inequalities in Health: An Analysis of Mortality Patterns Using Two Social Classifications." *Sociology of Health and Illness* 18:455–74.
- Breen, Richard, and John H. Goldthorpe. 1997. "Explaining Educational Differentials: Towards a Formal Rational Action Theory." *Rationality and Society* 9:275–305.
- Erikson, Robert, and John H. Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press.
- Evans, Geoffrey. 1992. "Testing the Validity of the Goldthorpe Class Schema." *European Sociological Review* 8:211–32.
- . 1993. "Class Prospects and the Life-Cycle: Explaining the Association between Class Position and Political Preferences." *Acta Sociologica* 36:263–76.
- Evans, Geoffrey, and Colin Mills. 1998. "Identifying Class Structures: A Latent Class Analysis of the Criterion-Related and Construct Validity of the Goldthorpe Class Schema." *European Sociological Review* 14:87–106.
- Goldthorpe, John H. 1996. "Class Analysis and the Reorientation of Class Theory: The Case of Persisting Differentials in Educational Attainment." *British Journal of Sociology* 45:481–506.
- . 1997a. "The 'Goldthorpe' Class Schema: Some Observations on Conceptual and Operational Issues in Relation to the ESRC Review of Governmental Social Classifications." Pp. 40–48 in *Constructing Classes: Towards a New Social Classification for the UK*, edited by David Rose and Karen O'Reilly. London: Office for National Statistics.
- . 1997b. "Social Class and the Differentiation of Employment Contracts." Paper presented at the European Consortium for Sociological Research Conference, Rational Action Theories in Social Analysis: Applications and New Developments. Stockholm, October.
- . 2000. *On Sociology: Numbers, Narratives and the Integration of Research and Theory*. Oxford: Clarendon Press.
- Korpi, Walter. 1983. *The Democratic Class Struggle*. London: Routledge.
- . 1985. "Economic Growth and the Welfare State: Leaky Bucket or Irrigation System?" *European Sociological Review* 2:97–118.
- Rieger, Elmar, and Stephan Leibfried. 1998. "Welfare State Limits to Globalization." *Politics and Society* 26:363–90.
- Rose, David, and Karen O'Reilly, eds. 1997. *Constructing Classes: Towards a New Social Classification for the UK*. London: Office for National Statistics.
- . 1998. *The ESRC Review of Governmental Social Classifications*. London: Office for National Statistics.
- Schumpeter, Joseph A. 1954. *History of Economic Analysis*. London: Allen & Unwin.

A Neo-Utilitarian Theory of Class?¹

Dietrich Rueschemeyer and James Mahoney
Brown University

Aage Sørensen aims to develop a structural theory of inequality that is equal in format and comprehensiveness to Marx's theory of class yet avoids the flaws of that theory, which derive from its grounding in the premises of classical economics. Applying neo-utilitarian theory to class analysis, Sørensen argues that access to enduring rents can inform a new conceptualization of "class as exploitation" and thereby put the sociological enterprise of class analysis on a sounder basis. The need for such conceptual reformulation grows out of the marginalist turn in economic theory and the related demise of the labor theory of value as well as, Sørensen suggests, the failure of subsequent class analysts to invent alternative theories of structural inequality in which exploitation generates antagonistic interests. Although many scholars—Marxist and non-Marxist—have proposed alternative class schemes, nearly all of these are based on an understanding of "class as life conditions." According to Sørensen, class as life conditions—defined by the total wealth controlled by similarly situated actors—does not necessarily create antagonistic interests and thus provides a poor foundation for understanding how class position generates mobilization and conflict.

We agree that the labor theory of value of classical economics is indefensible, leaving Marxist theory without its primary basis for identifying class exploitation.² However, we believe that Sørensen's alternative theory of rent suffers from its own serious problems and ultimately fails as a basis for a renewal of class analysis. The theory assumes that rent-based deviations from a fully competitive market can be used as a baseline for assessing the existence of structural inequality and exploitation. Yet, perfectly competitive markets as such carry no moral authority, and above-market returns in the form of rents cannot be morally condemned as unjust and exploiting simply because they deviate from the returns on assets in competitive markets. In fact, the marginalist theory of productivity remained

¹Direct correspondence to Dietrich Rueschemeyer, Department of Sociology, Maxcy Hall, Box 1916, Brown University, Providence, Rhode Island 02912. E-mail: Dietrich_Rueschemeyer@brown.edu

²For summaries of the limitations of the labor theory of value see Elster (1985, chap. 3) and Gordon (1990).

largely immune from the essentialist and normative interpretations that were uniformly associated with the labor theory of value, though there have been exceptions, both today and in the past.³ Sørensen himself is unprepared to defend the outright use of fully competitive markets as a moral yardstick, because, as he suggests, "nothing guarantees that efficient labor markets create good lives" (p. 1553).⁴ Yet, the very use of the word "exploitation" for enduring rents suggests implicit essentialist/normative assumptions about the benevolence of perfect market functioning, assumptions that underlie other formulations as well.⁵

For the effective empirical analysis of class, the moral judgments of social scientists are less important than the valuations of the populations they study. Yet, we know of no society in which all rent-based deviations from competitive market returns on assets are universally treated as unjust. Rather, in all real societies, deviations from perfectly competitive markets and their implications for moral judgment are extremely heterogeneous in character. Rents differ in the consequences they have in society, in the social reactions they engender, and—intertwined with their impact on social life—in the moral evaluation they receive. Thus, Sørensen's use of "exploitation" to characterize all rents represents either an esoteric normative judgment of neo-utilitarian thinkers, which is socially meaningless in real societies, or it is a misnamed concept without any moral significance.

Given the lack of moral grounding in Sørensen's theory of class exploitation, there are good reasons to believe that the advantages and disadvantages conferred by rents will not become crystallization points for class consciousness and collective action. This is true in part because moral evaluations—always implicitly or explicitly inspired by a vision of a good society—are, together with shared interests, an indispensable ingredient in broad-based political class formation. Furthermore, because different kinds of rents have multiple consequences and vary considerably in their overall social impact, it is not clear that rents as such, even as they are directly advantageous or disadvantageous for a given set of actors vis-

³ One such normative interpretation of the neoclassical concept of productivity occasioned passionately negative responses by Max Weber and others, giving rise to the famous debate about the role of values in modern social science (see Rueschemeyer and Van Rossem [1996, pp. 145–46, 147–48]).

⁴ See also his comment in an earlier paper that "these assumptions [about perfect competition] are useful for clarifying the arguments. They in no manner imply that such a society [in which all transactions take place in a perfect market] is to be preferred over other societies" (Sørensen [1996, p. 1345 n. 7]).

⁵ As in the following: "Competition in the labor market guaranties that the capitalist pays no more and no less than the worker contributes to his production" (p. 1531).

à-vis the rest of the society, will in their total set of direct and indirect consequences effectively define antagonistic interests. As a result, there is little reason for believing that rents will consistently generate class consciousness and collective organization.

These arguments—and their implications for ongoing work from the perspective of class as life conditions—require some elaboration.

RENTS, COLLECTIVE INTERESTS, AND CLASS FORMATION

The presence or absence of rents, and thus “exploitation,” is defined by Sørensen in relation to the ideal-type of perfect market functioning. In a fully competitive market, exploitation does not occur, since there are no owners of assets who receive payment above market value at the expense of nonowners. By contrast, in less than fully competitive markets, control over rent-producing assets privileges owners over nonowners and provides a latent basis for conflict, since, *ceteris paribus*, owners have an objective interest in preserving (and expanding) their rent, while nonowners have an interest in destroying it. The “*ceteris paribus*” clause is important, and we will suggest that it often cannot be sustained. Nevertheless, we note here the appeal of Sørensen’s framework: it avoids the problems of Marx’s labor theory of value while still seemingly identifying objective conflicts rooted in property rights, broadly conceived.

This framework makes sense only so long as rents have no significant social consequences beyond the zero-sum gains and losses they distribute to owners and nonowners. Yet, rents often create additional public advantages that have varied implications for actors situated in different places of the overall rent landscape. These social benefits may be highly valued by many actors, including those who are not owners of the rent in question. Whether or not the maintenance of such a rent is then in the “real” interests of broad societal populations, including those without ownership rights, depends on how the social benefits of the rent are evaluated in comparison to its deadweight costs. In actual societies where individuals are given the choice, they often decide that the social benefits outweigh the deadweight costs.

This suggests that certain kinds of rents will not generate antagonistic interests and conflict because—given their indirect as well as their direct consequences—they ultimately serve the interests of nonowners and owners alike. For example, rents from long-term patents give monopolistic advantages to the patent holders, but they also instigate innovation from which many others benefit. Likewise, the state creation of monopoly rents through licensing confers advantages to license holders, but such rents also provide public benefits such as competent doctors, safe cars, and

skilled teachers. Large populations often choose to endorse these rents because they lead to the delivery of goods and services that might otherwise not be available. Similarly, rents based on social legislation and labor market practices protect much of the population from poverty, which has desirable social consequences beyond those conferred to direct beneficiaries of the rent. Although Sørensen argues that "it is to the advantage of the capitalist class to produce a labor market conforming to the assumptions of neoclassical economics" (p. 1554), he is also well aware that the diminution of rents in labor legislation and their near-abolition in the labor market have been associated with expanded inequality and the deterioration of living conditions in the United States (pp. 1550–53). It is debatable whether the abolition of these rents—given both their direct and indirect social consequences—serves the collective interest of the capitalist class. Indeed, the dominant classes in other capitalist political economies see little advantage in similar market policies, opting instead for more comprehensive systems of social provision.

It thus makes little sense to generalize about all rents as representing a latent basis for class formation and mobilization. Rents that distribute advantages to groups that already enjoy privileged positions within society, while imposing losses on disadvantaged groups without providing any favorable social by-products, may provide one basis for the creation of class interests and class mobilization. Yet, many kinds of rents do not meet these conditions and thus will not foster direct conflicts of interest between rent owners and nonowners. For instance, we are skeptical that rents from minimum wage legislation will elicit the same value and interest-inspired response as will, for example, rents from particularistic tax exemptions that exclusively benefit the wealthy.

Even if some rents do generate antagonistic objective interests, we are skeptical that their critique will carry the moral authority to produce broad-based class formation and mobilization. Despite the limitations of the labor theory of value, Marx's understanding of that theory and his related vision of a socialist society characterized by abundance, equality, and democracy provided a normative benchmark in the name of which disadvantaged classes could mobilize. Sørensen's exploitation-free society marked by fully competitive markets offers no similar benchmark; rather, as Sørensen acknowledges, there is nothing inherently desirable about a society in which all transactions take place in fully competitive markets. Indeed, because most people in real societies do not consider the principle of rents to be unjust or morally indefensible, they are hardly likely to develop class consciousness and pursue collective action in the name of ending rents and establishing a more competitive market society.

The lack of moral authority underlying Sørensen's definition of exploitation leads to the identification of forms of "exploitation" that violate

commonsensical understandings. For example, much of his discussion treats economically marginal segments of the U.S. population as an "exploiting" class that benefits at the expense of more wealthy and powerful groups through control over rent-producing assets. Likewise, he argues that the *elimination* of "exploitation" through the destruction of rents in the labor market *increases* the marginalization of economically disadvantaged groups. Sørensen (p. 1531) himself rejects Roemer's (1982) structural theory of exploitation precisely because, as Roemer acknowledges, it can "happen that those who own very little of the means of production are exploiters and those who own a lot are exploited" (Roemer 1986, p. 262). The same problem arises with Sørensen's concept of class as exploitation, raising the issue of why his concept should be preserved and Roemer's abandoned.

In short, we believe there are two fundamental problems with Sørensen's argument about class as exploitation. First, given the heterogeneity of rents and their diverse social consequences, there is little reason to suppose that all rents will necessarily generate antagonistic interests. Second, because a pure market model lacks moral authority, it is not clear on what normative grounds similarly situated actors will come to perceive rents received by others as harmful and unjust exploitation and as the occasion for collective action.

CLASS AS LIFE CONDITIONS AND POLITICAL CLASS MOBILIZATION

In contrast to Sørensen's understanding of class as exploitation, his concept of class as life conditions—which we would define roughly in Weberian terms as "life chances" in asset markets along with a degree of social closure—will often produce shared ideas and attitudes, which, in turn, can lead to the pursuit of collective goals through organized class action. Sørensen downplays the value of such an approach because it fails to identify mechanisms through which the latent objective interests of classes come into conflict with one another. Yet, assumed "objective" or "real" interests are always "essentially contested" and ultimately raise questions of a moral and political character (Lukes 1974, p. 34). For the purposes of empirical analysis, class theories based on such antagonistic interests are useful only insofar as they can be linked to observable patterns of class consciousness and collective organization. If such systematic effects are absent—and we have argued that this is the case with Sørensen's rent theory of class—there is little reason to prefer a theory of class as exploitation over a theory of class as life conditions.

Sørensen points out that "there is an abundance of research that shows class as life condition indeed is a powerful determinant of all kinds of

outcomes" (p. 1538). We would simply add that included in these outcomes are forms of collective action in which individuals self-consciously identify themselves as part of a class and carry out political mobilization on that basis. Sørensen suggests that life-opportunity classes can produce important outcomes because individuals orient their behavior to long-term living conditions that can be expected from their wealth profile. However, as Weber (1978, pp. 927–30) noted, while differential access to goods and opportunities for income is a defining component of economic class, it does not ensure that individuals of a given class will share similar preferences, much less become a collective actor in history. Even if classes do have an objective interest in collective action (a question that can only partially be decided through empirical evidence), they may not mobilize on behalf of this interest. Hence, for empirical analysis, the critical question is not only one of identifying the interests of classes, but of understanding the processes through which such interests are socially constructed.

At present, our understanding of the social construction of class interests remains fragmented at best, but such an approach is still the most promising avenue for class analysis today and likely for a long time to come. We now have some knowledge about the social processes through which life condition classes are transformed into collective actors. For example, we know that individuals in particular classes are often exposed to common organizational processes that construct similar interests and provide opportunities for collective action. Sørensen rightly urges us to link the study of class formation to the "rich literature on social movements that addresses the problem of when interests will effectively be translated into action, emphasizing resource mobilization, political processes, and the collective action problem" (p. 1545). Likewise, we have tentative support for several hypotheses about the effects of geographic patterns, cultural traditions, and the structure of state action and politics on the formation of classes (see, e.g., Rueschemeyer, Stephens, and Stephens 1993, pp. 53–57). Future progress on this research—as with progress throughout the social sciences—will likely continue to take place through incremental steps rather than leaps and bounds.

We believe Sørensen's rent theory of class may have something to offer to this research. To realize this potential, however, it is necessary to abandon the idea that rents are a universal basis for exploitation and instead develop more specific hypotheses about the conditions under which actors come to perceive rents as exploitative. The visibility of the rent in question and the extent of costs and benefits that it distributes will obviously be relevant. But the nature of the winners and losers of rents must also be taken into consideration. For instance, rents that overlap with life condition classes in ways that distribute benefits to advantaged classes and costs to subordinate classes, without providing tangible additional social bene-

fits to the losers, are more likely to form a basis for political mobilization than other kinds of rents. Distinguishing these and other kinds of rents on the basis of their potential for mobilization in real societies could represent a valuable step forward for class analysis.

We argue, then, that there are good reasons for believing what Sørensen calls "class as life conditions" will identify sites of mobilization and conflict. Indeed, people are more likely to mobilize around life conditions than around "exploitation"—if exploitation is defined in terms of access to rent producing assets. As a result, if Sørensen's proposed framework is to be useful for empirical analysis, it must be redirected toward efforts to identify the specific conditions under which particular kinds of rents will lead life condition classes to construct class interests and pursue political class mobilization.

Sørensen rightly calls for finding causal mechanisms that identify antagonistic interests and explain political class formation. Yet, his neo-utilitarian project offers—much like its rational choice siblings—a comprehensive promise of theory advancement that is unlikely to be fulfilled. For the foreseeable future the best we can hope for is to identify causal mechanisms of a limited reach, in which the conditions under which they hold are not fully specified—"bits of sometimes true theory," as Arthur Stinchcombe puts it.⁶ Although this research program is modest, we believe it has more to offer than neo-utilitarian schemes that make promises that cannot be delivered.

A CONCLUDING REFLECTION ON METATHEORY

Sørensen's proposal is an exercise in metatheory or, as Robert Merton called it, "general sociological orientations" (Merton 1957, pp. 87–89). Metatheories cannot be tested directly. They point to problems, identify relevant factors and variables, and offer arguments why these factors and variables are pertinent. They aid in the formulation of testable hypotheses, but are themselves not judged as true or false but rather as useful or useless in the construction of empirical theory. We have expressed our skepticism that Sørensen's conception of class as exploitation amounts to a useful reorientation of class theory.

Nearly uniformly, however, authors and readers are interested in metatheories for reasons that go beyond their utility in empirical theory construction. How else can one explain the intense investment scholars display when they discuss the role of culture in sociological explanations or

⁶ See Stinchcombe (1991, p. 31; 1993, p. 267); he refers to Coleman (1964, pp. 516–19) as the inventor of the phrase.

the overall explanatory power of rational choice models? Metatheories tend to have a dual face. The same ideas that in one sense are tools of empirical theory construction also function as building blocks for broad worldviews, for more or less coherent ideological and philosophical orientations. The label "grand theory of society" points to such a fusion of a sociological orientation in the narrow and technical sense with a comprehensive interpretation of human social life.

Sørensen's proposal for a reconceptualization of class theory is not simply a technical tool for generating empirical hypotheses. It, too, evokes ideological-philosophical implications as it makes the perfectly functioning competitive market the criterion for defining "exploitation." As we noted earlier, Sørensen rejects an explicit commitment to the neo-utilitarian ideal of a society thoroughly shaped by competitive market transactions, but, as it stands, his proposal still entails similar premises.

The tendency of metatheoretical arguments to acquire this dual character requires that authors and readers take responsibility for both sides. This is not a demand for one version or another of political correctness. The politically incorrect model may generate strong and otherwise not easily attainable empirical insights, and the politically correct orientation may conceal what Max Weber called "inconvenient facts." Thorough separation of the two sides is one reasonable response. Explicitly taking positions on both sides while differentiating clearly between orientational guesses and philosophical and ideological commitment is another.

REFERENCES

- Coleman, James S. 1964. *Introduction to Mathematical Sociology*. New York: Free Press of Glencoe.
- Elster, Jon. 1985. *Making Sense of Marx*. Cambridge: Cambridge University Press.
- Gordon, David. 1990. *Resurrecting Marx: The Analytical Marxists on Freedom, Exploitation, and Justice*. New Brunswick, N.J.: Transaction Books.
- Lukes, Steven. 1974. *Power: A Radical View*. London: Macmillan.
- Merton, Robert K. 1957. *Social Theory and Social Structure*, rev. ed. Glencoe, Ill.: Free Press.
- Roemer, John. 1982. *A General Theory of Exploitation and Class*. Cambridge: Cambridge University Press.
- . 1986. "Should Marxists Be Interested in Exploitation?" Pp. 260–82 in *Analytical Marxism*, edited by John Roemer. New York: Cambridge University Press.
- Rueschemeyer, Dietrich, Evelyne Huber Stephens, and John D. Stephens. 1992. *Capitalist Development and Democracy*. Chicago: University of Chicago Press.
- Rueschemeyer, Dietrich, and Ronan Van Rossem. 1996. "The Verein für Sozialpolitik and the Fabian Society: A Study in the Sociology of Policy-Relevant Knowledge." Pp. 90–153 in *States, Social Knowledge, and the Origins of Modern Social Policies*, edited by Dietrich Rueschemeyer and Theda Skocpol. Princeton, N.J.: Princeton University Press.
- Sørensen, Aage. 1996. "The Structural Basis of Social Inequality." *American Journal of Sociology* 101:1333–65.

- Stinchcombe, Arthur L. 1991. "The Conditions of Fruitfulness of Theorizing about Mechanisms in Social Science." Pp. 23–41 in *Social Theory and Social Policy: Essays in Honor of James S. Coleman*, edited by Aage Sørensen and Seymour Spilerman. Westport, Conn.: Praeger.
- . 1993. "Monopolistic Competition as a Mechanism: Corporations, Universities, and Nation-States in Competitive Fields." Pp. 267–305 in *Social Mechanisms: An Analytical Approach to Social Theory*, edited by Peter Hedström and Richard Swedberg. Cambridge: Cambridge University Press.
- Weber, Max. 1978. *Economy and Society*, 2 vols. Edited by Guenther Roth and Claus Wittich. Berkeley: University of California Press.

NOTE.—Illness prevents Aage Sørensen from replying here. His response to the commentators will appear in a future issue.

State-Society Relations and the Discourses and Activities of the 1989 Beijing Student Movement¹

Dingxin Zhao

University of Chicago

Many writings emphasize the importance of Chinese culture to the 1989 Beijing Student Movement. Comparing the 1989 movement to the May 4th movement of 1919 and to the December 9th movement of 1935–36, this study finds the rhetoric and activity patterns of the 1989 movement to be actually more traditional than those of its fore-runners. This is quite a paradox, considering the scale of social change in China during the 20th century. This article argues that the 1989 movement's traditionalist outlook can be explained by three structural conditions, all involving state-society relations. In comparison with the states of the earlier movements, the state during the 1980s had a higher capacity for repression. To avoid immediate repression, students tended to hide their real demands and goals behind safer and culturally congenial forms of action. Second, China in the 1980s had comparatively weaker independent civil organizations. Hence, the movement was poorly organized, and competing activities frequently occurred in the same time and location. What determined the efficacy of these activities then were less the actors' intentions than the perception of those who were observing the movement. This study argues that people's acceptance of particular movement activities is shaped by their perception of state legitimization. During the 1980s, most Chinese saw that the state's legitimacy lay in its moral performance, close to what was in traditional China. Therefore, people were more receptive to culturally and morally charged movement activities. This furthered the domination of traditionalism during the 1989 movement.

Because of its scale, duration, tragic ending, and impact on current Chinese politics, the 1989 Beijing Student Movement (BSM) attracts a great deal of scholarly attention. To date, scholars have tried to explain the

¹ The author would like to thank William Parish, Roger Gould, Mayer N. Zald, John A. Hall, Qin Chen, Samuel Baker, Harris Kim, and Jennifer Pashup for their helpful comments. This research was supported by the J. David Greenstone Memorial Fund. Direct all correspondence to Dingxin Zhao, Department of Sociology, University of Chicago, Chicago, Illinois 60637. E-mail: dzhao@mailhub.src.uchicago.edu

causes of the movement and its development by tracing such factors as the rise of civil society (Burns 1989; Huan 1989; Strand 1990; Sullivan 1990), the lack of civil society (Zhou 1993), the political opportunities provided by factionalism among top leaders (Cheng 1990; Dittmer 1990; Kristof 1990; Nathan 1990; Tsao 1991), the influence of international media (He and Zhu 1994), the importance of ecology-based mobilization (Zhao 1998), and the decline of student control systems in Chinese universities during the 1980s (Rosen 1991; Zhao 1997).

The preponderance of writings on the topic, however, emphasizes the importance of Chinese culture (and the accompanying emotional elements) to the 1989 movement (Wasserstrom and Perry 1992). During the movement, students often set picket lines to prevent other elements of the Beijing population from joining their demonstrations. This type of activity is interpreted as reflecting a sense of elitism, a sentiment deeply rooted in Chinese intellectual culture. It is also considered to be an important factor contributing to the failure of the movement (Chan and Unger 1990; MaCartney 1990; Perry and Fuller 1991). Students tended to march together by school, class, and often major. Calhoun (1994) argues that this manifested Chinese cultural values of solidarity, loyalty, and friendship. It is a fact that conflicts between students and the government had continuously escalated until the military crackdown. According to Pye (1990), this was because Chinese cultural sanctions on self-interested behavior led students (and the government) to attack the other party with lofty, symbolic, and moralistic rhetoric and activities. Since battles of morality and shame often have a positive feedback nature, such interactions portended a bloody ending. All in all, for many cultural scholars, the 1989 movement was a "street theatre" with Chinese culture supplying the "scripts" (Esherick and Wasserstrom 1990; Wasserstrom 1991).²

Obviously, this type of analysis has weaknesses (Mann 1993, chap. 6; McAdam, Tarrow, Tilly 1997; Tarrow 1998). Logically, to attribute cultural manifestations to culture courts tautology. Empirically, such analysis tends to overemphasize cultural stability and to assume that the relationship between culture and behavior is deterministic. It usually fails to explain questions such as Why do waves of movements emerge in some periods and not in others, and why are some more adept at manipulating cultural symbols than others? Indeed, most of these works tell stories in which culture only functions as a "metanarrative" (Tarrow 1998, pp. 17–18).

² This research tradition is influenced by discourse analysis (Jabri 1996; Steinberg 1994; Terdiman 1986), that of poststructuralists (e.g., Derrida 1978, 1981; Foucault 1980, 1994) as well as interpretative anthropologists (Geertz 1973, 1983). Many new generation French Revolution scholars also follow this tradition (Baker 1990; Furet 1981; Hunt 1984; Sewell 1985).

The cultural perspective often fails to explain a reasonable amount of variation of the 1989 BSM. Let me provide two examples. In the beginning of the 1989 BSM, most students acted very cautiously. However, by early June, many were risking their lives to resist the military repression. Calhoun (1991, 1994) argues that the students became braver because they became increasingly committed to a culturally embedded identity during their movement participation. Yet, as a matter of fact, by the end of May, most Beijing students were very tired and went home; the Tiananmen Square occupation was sustained by students who were continuously arriving in Beijing from other provinces after mid-May. Since identity cannot be borrowed from someone else, we cannot attribute students' bravery during the military repression to an internal identity transformation. Following the logic that battles of morality and shame would only escalate confrontations, Pye (1990) predicted that, after the repression, the government would take severe revenge on students and reformers and that China's reform would be stopped. Yet, Chinese politics in the 1990s was not governed by a desire for revenge, as Pye predicted, and China's economic reform continues.

Although cultural analysis is mechanical, many of its observations are not without foundation. People can be mobilized by cultural symbols and norms as well as by material interests. Social movements and revolutions all have important discursive dimensions (Hunt 1984; Jabri 1996; Sewell 1985; Steinberg 1994). Empirically, it is undeniable that the languages and activities that students used during the 1989 BSM were deeply rooted in Chinese culture and had a significant impact on the movement. The challenge is how to analyze this impact in a causal manner. Since comparison is the essential method of falsification in the empirical sciences, I compared the movement rhetoric and activities of the 1989 BSM with those of the May 4th movement of 1919 and the December 9th movement of 1935–36—the other two large student movements in Chinese history. My initial hypothesis was that if the patterns of a social movement are determined by culture, the 1989 BSM should appear less traditional than the earlier movements, since Chinese culture has become significantly more modern during the 20th century.

To my surprise, I found that the rhetoric and activity patterns of the 1989 BSM were actually truer to traditional Chinese culture than those of the two earlier student movements. For instance, during the 1989 BSM, Beijing students tried to exclude Beijing workers from participating in the movement, an action that, as has been suggested (Chan and Unger 1990; McCartney 1990; Perry and Fuller 1991), was motivated by a traditional sense of elite status. In contrast, during the May 4th movement, students praised workers as sacred (*laogong shensheng*), tried to earn their own living through labor (*bangong bandu*), and mobilized workers and

merchants to form a broad coalition. During the December 9th movement, students launched a rural crusade to the south to mobilize China's peasantry into an anti-Japanese coalition.³ These findings posed a puzzle that pushed me to consider other differences behind the three student movements.

I found that the 1989 BSM was developed under different "state-society relations" than had existed during the two earlier student movements. Unlike the eras of the earlier movements, the 1980s found China with a unitary state, weak independent organizations, and state legitimacy based on moral and economic performance. My major argument is that because this unitary state had allowed state authoritarianism to penetrate deeply into society, students in the 1980s had a stronger fear of repression than students during the earlier movements. Some culturally interpretable patterns of movement activities, such as the picket lines, were strategies intended to avoid immediate repression. Moreover, because of the weakness of associational life, many activities during the 1989 BSM were initiated by individuals or a small group of students without preplanning, with multiple activities often occurring simultaneously in a given time and space. These activities might be rooted in Chinese culture, in a communist style of mobilization, or in Western influences; people also acted from any number of motives. The simultaneous occurrence of several activities in the same location made the psychology of sympathetic bystanders an important factor in shaping the patterns of the movement. One of the central arguments of this article is that people's views on the proper behavior of movement activists and the government were shaped by their understanding of state legitimation. During the late 1980s, the notions of most urban Chinese of what legitimates a state had changed from communist ideology to moral performance, a notion more in accordance with the ideas found in an earlier, more traditional China. Those who watched the movement were thus more receptive to culturally and morally charged rhetoric and activities and became very emotional when the state reacted harshly to morally charged acts. This further intensified the traditionalism expressed during the 1989 BSM.

THE STATE-SOCIETY RELATION AND ITS ESSENTIAL DIMENSIONS

The phrases "state and society" and "state-society relations" existed only marginally in sociology before the 1970s (e.g., Bendix 1968; Blalock 1969; Morley 1949, chap. 5). Back then, the dominant theories and research agendas in sociology were largely society centered (Skocpol 1985); "state-society relations" was more a casual phrase than a theoretical framework.

³ See Israel (1966, pp. 134–38) for more details about the event.

The situation changed between the late 1970s and the 1980s. One of the pioneering works was Skocpol's (1979) classic study on the role of the state in great social revolutions. Skocpol (1979, p. 27) treats the state "as an autonomous structure—a structure with a logic and interests of its own not necessarily equivalent to, or fused with, the interests of the dominant class in society or the full set of member groups in the polity." Since then, state-centered research has blossomed not only in the study of social revolutions (Foran 1997; Goodwin 1989; Goodwin and Skocpol 1989; McDaniel 1988, 1991; Wickham-Crowley 1992) but also in studies of other types of contentious politics, such as working-class movements (Katznelson and Zolberg 1986; Lipset 1983; Mann 1993, chaps. 15–18; Marks 1989) and contemporary social movements (Jenkins and Klandermans 1995; Kitschelt 1986; Kriesi 1996; Kriesi et al. 1995), and in economic development (Amsden 1989; Evans 1995; Gold 1986; Haggard 1990; Wade 1990).

As this research has developed, more and more societal factors have been brought back into the state-centered analysis. In the area of economic development, instead of arguing for a "strong state," Zhao and Hall (1994) and Evans (1995) try to capture the importance of state-society linkages in economic development with concepts such as "bounded autonomy" or "embedded autonomy." In the study of social revolutions, scholars including Goldstone (1991) and McDaniel (1991) emphasize not only the structure and nature of the state, but also factors such as population density, economic structure, and ideational factors. By far the largest amount of recent research focusing on state-society relations concerns democratic transition and consolidation (Casper and Taylor 1996; Chehabi and Stepan 1995; Diamond, Linz, and Lipset 1988; Linz and Stepan 1996; Schmitter 1993; Stepan 1978, 1989; Weil 1996). Although their approaches vary, most studies of this type treat democratization as an interactive process between the state and society, examining factors such as the nature of the state, the nature and strength of civil society, and socioeconomic structure. By the mid-1990s, scholars generally accepted that the state and society transform each other through a process with outcomes that do not always reflect the aims or will of either (Migdal 1994, p. 23). Obviously, the state-society relation perspective is intended to achieve a more balanced understanding of some political processes by apprehending not only the structure and nature of the state, but also the interactions between the state and society.

China has the longest uninterrupted state tradition in world history. Since the communists took power, Chinese government has initiated numerous social programs aimed at changing the state, society, and their relations. Each of these programs had a profound impact on the population. The Great Leap Forward, the Cultural Revolution, and, to a great extent, the current reform are only the best-known examples. In fact, for those who know China, it is very hard not to think of Chinese politics in

terms of the state and of state-society relations. I imagine that it is for this reason that many China scholars have adopted approaches that either directly use or imply the idea of state-society relations.

In analyzing the 1989 BSM, Tsou (1991) uses the concept of state-society relations. Zhou's (1993) "large number" and "unorganized interests" concepts also contain the idea of state-society relations. In other research areas, Shue (1988) discusses the nature and limits of the state capacity of the communist government, and Wong (1997) theorizes on the general patterns of state-society relations in both historical and contemporary China, through a comparison with the European experience. Beyond these, Tang and Parish's (2000) recent study on social changes during Deng's era, Shirk's (1993) interpretation of the drives of China's economic reform, Liu's (1996) study on China's mass politics, Zhou's (1996) work on the role of Chinese farmers during the early economic reform, and Yang's (1996) research on the impact of the great famine on the later state policies, to mention just a few works published in the 1990s, have all consciously or unconsciously adopted a state-society relation perspective.

However, of this plethora of studies, few focus on the impact of state-society relations on the discursive and symbolic dimensions of political processes. Moreover, the concept of state-society relations is capacious. Most studies, with the noticeable exception of Linz and his associates' work on democratic transition and consolidation, have used it as a general concept under which quite a few meanings are attached but not clearly specified. Although a complete operationalization of the concept into several variables is neither possible nor necessary, we should at least turn this general concept into a powerful research perspective by specifying basic dimensions that should be attended to in most research. To this end, I suggest three impure but nonreducible dimensions of state-society relations: the nature of the state, the nature of society, and the links between the state and society. Later on, I will define and discuss the basic range of variation of each of these three dimensions, with regard to the three student movements I am comparing.

After a short methodological note and a section that introduces the three student movements, the rest of this article proceeds as follows. First, I delineate the patterns of movement rhetoric and activities during the 1989 BSM (the variations of dependent factors). Second, I define the three dimensions of state-society relations that I use to explain the important variations elaborated in the previous section (the independent factors). Third, I describe the different state-society relations behind the May 4th, the December 9th, and the 1989 movements along the three identified dimensions (the variations of independent factors). Finally, I analyze how different state-society relations shaped movement rhetoric and activities during the 1989 BSM (the causal relations between the dependent and independent

factors). For those not familiar with the comparative-historical approach, I have indicated the logic of each section parenthetically by quantitative analogies. These remain just analogies, however, because the structural dimensions that I have adopted are impure constructions rather than fully operationalized variables and the discourses and activities that my analyses focus on are not dependent variables but impure ramifications of the state-society interactions. Due to the nature of the work, a full elaboration of the causal logic of this article is not given until the last section.

A METHODOLOGICAL NOTE

This article compares movement activities in the May 4th movement of 1919, the December 9th movement of 1935–36, and the 1989 BSM. These are all large-scale student movements that took place in urban China. Thus, some undesirable variations are “controlled.” The main factors remaining uncontrolled include the international environment, learning, and knowledge accumulation processes from previous collective action experiences; the goals of each movement; and relationships between the state, students, and the general public.

A major part of the data on the May 4th and the December 9th movements was obtained from published monographs, witness accounts, and memoirs. I did independent research in Beijing and Shanghai’s libraries and archives in 1997. The data for the 1989 BSM came mainly from my own interviews. A total of 70 people were interviewed between 1992 and 1993, 40 in Beijing and 30 in Montreal, Canada. Of the 70 informants, 48 were university students during the movement, 16 were political control cadres or teachers in Beijing universities, and 6 were young researchers from various cultural and academic institutions. I have given the details of the interview methodology elsewhere (see Zhao 1996, 1997, 1998).

To limit the length of this article, I adopt an unbalanced comparison. My major focus is on the 1989 BSM. The two earlier student movements are used mainly as contrast cases to the 1989 BSM. Discussions on the two earlier movements will be kept at a reasonably minimum level. For more on the earlier movements, see Chow (1967) and Israel (1966).

THE THREE STUDENT MOVEMENTS

The May 4th movement of 1919, the December 9th movement of 1935–36, and the 1989 movement are the three largest student movements in 20th-century China. They all started in Beijing and spread across the nation. Here I only briefly introduce these events.⁴ The May 4th movement

⁴ For interested readers, Chow (1967), Yeh (1996), Israel (1966), Calhoun (1994), and Lin (1992) offer case studies.

was the first major student movement in modern China. Since 1917, China's new intellectuals and students initiated a vast modernization drive that aimed to strengthen the country through science and democracy. This modernization drive to a great extent paved the ways to the rise of the movement. The movement, however, was named after the May 4th incident in 1919: on that day, over 3,000 Beijing students marched in the streets in an anti-Japanese demonstration protesting the Shandong Resolution of the Versailles peace conference. Joined by merchants, workers, and other urban residents, anti-Japanese protests soon spread to many other Chinese cities. Eventually, the government of Northern China had to dismiss three pro-Japanese officers and to refuse to sign the Versailles peace treaty. The impact of the May 4th movement was profound. Politically, it contributed to the rise of the Chinese Communist Party (CCP) and the reorganization of the Nationalist Party (Guomindang). Culturally, it facilitated the rise and dominance of vernacular literature and mass education and the decline of Confucianism and traditional ethics.

While the May 4th movement was at once a prodemocratic and a nationalistic movement, the December 9th movement was mainly a nationalistic reaction to Japanese aggression. By 1935, Japan had occupied the whole Dongbei (Manchuria) region and a large part of the Hebei provinces. Beijing was virtually encircled by Japanese troops. On December 9, 1935, thousands of Beijing students marched to urge the government to resist the Japanese invasion. The movement spread to most of urban China in 1936. The impact of this movement was also monumental. It helped the CCP rebuild their urban base, which had been destroyed by the Guomindang. It also split the governing elites and induced the Xi'an Incident.⁵ A series of chain reactions during and after the movement led to the establishment of a united front between Guomindang and the CCP to resist the Japanese invasion, an event that paved the road for the communist victory in China (Johnson 1962).

The 1989 BSM was prodemocratic, yet it was also a direct reaction to emerging social problems such as high inflation, rampant official corruption, and the decreasing economic status of intellectuals. Beijing students started the movement on April 15, 1989, upon Hu Yaobang's sudden death.⁶ The movement achieved some success by early May. However, unlike the two earlier student movements, the 1989 BSM was not able to

⁵ On December 12, 1936, general Zhang Xueliang detained Chiang Kai-shek, the generalissimo, to try to force him to stop the war with the communists and lead the whole of China to resist the Japanese invasion. This is known as the Xi'an Incident.

⁶ Hu Yaobang was the general secretary of the CCP between 1981 and 1987. Forced to resign his position in part because of his liberal attitude toward student unrest in 1986, Hu became a respected figure among students and intellectuals.

sustain its success. While the government was pushed to make limited concessions, a few students started a hunger strike at Tiananmen Square on May 13 to demand more radical changes. The action drew hundreds of thousands of sympathizers once the condition of the hunger strikers deteriorated. The hunger strike was a big success in mass mobilization. However, it interrupted the Sino-Soviet summit and antagonized most of China's top state leaders. Thus, the government announced martial law. Yet, martial law could not be carried out immediately. On the night of May 19, in a popular belief that the soldiers were going to harm the students in Tiananmen Square, Beijing residents went out by the hundreds of thousands and stopped the army. The troops had to withdraw, and the student occupation of Tiananmen Square was preserved. Then, on June 3, the army entered Beijing once again. They met with violent resistance. Yet with much more resolute orders from the government, the troops pushed their way through the square, leaving behind several hundred dead and thousands more wounded. The movement was suppressed, but people did not forget. Much of Chinese politics since then has centered on the ghost of the movement and its aftermath. The impact of this movement on China's politics has yet to be fully realized.

PATTERNS OF MOVEMENT ACTIVITIES DURING THE 1989 BSM

This study investigates the impact of state-society relations on the patterns of social movement activities. However, I do not claim that social movement activities are completely shaped by state-society relations. In comparison with the May 4th movement of 1919 and December 9th movement of 1935–36, the 1989 BSM has three distinctive features. The movement was much more pro-Western in appearance. It bore a strong imprint of the Cultural Revolution and styles of communist mass mobilization. Finally, culturally embedded rhetoric and movement activities figured much more centrally in the 1989 BSM than in the two earlier student movements. Only the last pattern was closely related to state-society relations in China during the 1980s. Before examining it, however, I will briefly introduce each of the three features.

Pro-Western Characteristics of the Movement

During the 1989 BSM, students frequently flashed V-signs to indicate victory, built a Statue of Liberty turned "Goddess of Democracy" in Tiananmen Square, held countless press conferences to attract Western attention, provided Western and Hong Kong journalists with easy access to their headquarters, and routinely listened to foreign broadcast for news and

feedback.⁷ The list of pro-Western attributes can be expanded. The movement's pro-Western nature was also reflected in its language. English banners such as *For the People, By the People, Absolute Power, Absolute Corruption*,⁸ *Glasnost*, and *People Power* appeared everywhere.⁹ This pro-Western sentiment was new to China. Although the two earlier movements were anti-Japanese, students also expressed grievances with other foreigners. For example, during the May 4th movement, when Shanghai students demonstrated around Waitan Park in the British concession, several students smashed a board in front of the park which read *No Chinese and Dogs* (Xu 1987, p. 76). During the May 4th movement, the Guomindang elite had to make efforts to keep the movement from becoming explicitly anti-Western (Liu 1990).

Both the May 4th and the December 9th movements took place between the world wars. During that period, colonialism still dominated international geopolitics. Weak nations such as China were constantly bullied by big powers acting on their conflicting geopolitical interests. Given this situation, nationalism became a major feature of the movements. The May 4th students advocated Westernization mainly because they believed that democracy could save China from foreign aggression. By the time of the 1989 BSM, however, colonialism was seen as part of the distant past; intellectuals and students were no longer concerned about saving China. Instead they felt great pain as a consequence of the strong state that their predecessors had dreamed of and fought for. They also witnessed the decline of the Soviet Union and a new wave of worldwide democratization. State socialism was no longer treated as a viable model; capitalism and democracy had again become the engine of the world system. The pro-Western nature of the movement revealed fundamental changes in the international environment and the mood of Chinese intellectuals and students.

Imprint of the Communist Mass Mobilization

Communist education imprinted on the students a set of habits and iconic images. For example, during the 1989 BSM, students sang "The Interna-

⁷ According to McCartney (1990, pp. 9–10), "Access to inner sanctums [at Tiananmen Square] was granted only to the Western and Hong Kong press." Moreover, student leader Wang Dan visited Hong Kong reporters almost daily "to feel out their attitude to the students, to find out whether the outside world was interested in the movement."

⁸ See the photos in Xuese de Liming (1989, pp. 48, 68).

⁹ During the May 4th and the December 9th movements, students seldom used Western languages to express their demands. Even when students wrote their banners in English, they only intended to explain to the Westerners that their movements were not radical and anti-Western. See, e.g., Wasserstrom (1991, photos between pp. 124–25).

tionale" whenever they felt that their action was heroic and tragic. Although "The Internationale" is part of a legitimate set of symbols under the regime, students sang the song frequently because the song is rebellious in spirit and singing it was a standard way to express the type of emotions that students had learned through revolutionary dramas and films. However, some movement tactics popularized during the Cultural Revolution had an even greater impact on the 1989 BSM.

The hunger strike was arguably the single most important form of protest during the 1989 BSM. The strike led to the success of the movement, both in terms of the scale of mobilization it achieved and the worldwide attention it generated. It also sowed the seeds for martial law and the final military crackdown. Significantly, the tactic was not used during the May 4th and the December 9th movements but was popularized only during the Cultural Revolution. The impact of the Cultural Revolution on the 1989 BSM can be seen from Zheng Yi's account of his involvement in the movement (Zheng 1993, pp. 62–68). In his book, Zheng describes organizing a large-scale hunger strike during the Cultural Revolution. He also depicts how he passed on his experiences in that regard to activists at Beijing University, suggesting that they stage a hunger strike as early as April 22. Zheng Yi's fascinating story certainly had its impact on young listeners such as Wang Dan, Chai Ling, and Zhang Boli, all of whom were key figures in the hunger strike.

Between May 16 and 26 in 1989, around 172,000 students from various provinces went to Beijing by train (Wu 1989, p. 474). Students also went by other means of transportation. Thus, in late May, Tiananmen Square was filled by energetic newcomers who were reluctant to leave. These students went to Beijing not just to support the movement, but to take advantage of the opportunity to get a free trip to visit Beijing. This phenomenon, which had a tremendous impact on the development of the movement, is called *chuanlian*. To travel a long distance as part of the petition process started during the earlier student movements (Wasserstrom 1991, p. 88), yet it is the Cultural Revolution that gave *chuanlian* a new meaning—*chuanlian* as massive free tourism. By 1989, most old Red Guards regarded the Cultural Revolution as a personal as well as a collective tragedy. Nevertheless, most of them had good memories of *chuanlian* (Jin 1996). In the interviews, all the students whom I asked about *chuanlian* knew the story very well. Past novelties became this movement's routine.

The Traditionalism of the 1989 Movement

What is most stunning and hard to explain is that the language and activity of the 1989 BSM were actually more traditional and moralistic than

those of the May 4th movement of 1919 and the December 9th movement of 1935–36. This is not to suggest that China's traditional culture had no impact on the two earlier movements or that there are no similarities among the three movements. For example, in premodern China, kneeling in a public place was a major way by which the Chinese proclaimed their discontent and appealed to the public for support. Indeed, we see that during the May 4th movement students knelt in front of workers and merchants to try to mobilize them. Nevertheless, kneeling had never been as significant a movement tactic as it was during Hu Yaobang's funeral on April 22, 1989. Never before did kneeling bring out the emotions of hundreds of thousand of students and trigger a Beijing-wide class boycott. The same was true of staging protests during funerals. During the May 4th movement, the funeral protests at Tianjin and Shanghai were only minor activities (Chow 1967, pp. 129, 143). During the December 9th movement, Beijing students did hold a memorial service and demonstrated for Guo Qing, a high school student who had died in prison. However, when the government repressed this activity, no further radical actions followed (Han 1986, pp. 103–8; Israel 1966, p. 145). The 1989 BSM, by contrast, started after Hu Yaobang's sudden death and for its first week centered on mourning activities.

Students of the 1980s also brought back some of the rhetoric that their predecessors during the May 4th era had vehemently attacked and which had gradually disappeared from China's political discourse. Traditional Chinese virtues such as loyalty, filial piety, and images of extended family were under severe attack during the May 4th movement (Chow 1967, pp. 58–59); yet during the 1989 BSM, students and Beijing residents frequently used language drawing on these elements to mobilize emotions. One frequently encountered slogans that centered on extended family relations, such as: "Mama, we are not wrong!" "Mama: I am hungry, but I won't eat!" "Grandpa Zhao, uncle Li, come to save our big brothers and sisters!" and "Your big brothers have been very anxious!" In addition, one also encountered language that emphasized very traditional Chinese values: "To fast to death is only a small deed!"¹⁰ and "Loyalty and filial piety cannot be obtained at the same time."

Traditional language was not just an individual initiative. For example, the Hunger Strike Declaration, arguably one of the most important student declarations of the 1989 BSM, was also filled with traditional rhetoric and with elitist sentiment such as "When a person is about to pass away,

¹⁰ This is a slightly modified version of half of a traditional Chinese couplet, the other half of which goes: "Keeping up one's chastity is the most important matter." The couplet was used to teach women in traditional China to maintain chastity in widowhood.

he speaks kindly and wisely; when a bird is about to die, it cries sadly" and "Please forgive us. Your children cannot be loyal [to the country] and show filial piety at the same time." In comparison, few declarations during the May 4th and the December 9th movements bore traits of this sort; such language (and the values embedded in it) was in fact a target of attack in some student manifestos during the May 4th movement.

No wonder China scholars have been so committed to explaining the 1989 BSM by different facets of traditional Chinese culture. Yet while it is not difficult to find a strong empirical support for this connection, such explanations do not provide any clue as to why the 1989 BSM was more traditional and culturally driven than the May 4th movement of 1919 and the December 9th movement of 1935–36.¹¹

It is quite tempting to explain this difference in terms of the core issues of each of the three movements. This works partially for the May 4th movement. There is no doubt that the iconoclastic nature of the movement delegitimized the use of traditional rhetoric and tactics in the movement. However, this explanation does not work for the December 9th movement, because it was almost purely nationalistic. There is little reason to explain why a nationalistic movement in the 1930s was less traditional than a prodemocratic movement in the 1980s.

One might also argue that activity patterns of the 1989 BSM point to the resilience of culture. Yet, it is illogical to assume that youths of the 1980s were more exposed to traditional culture than the May 4th and the December 9th youths. In the 20th century, China has arguably undergone the most dramatic social change in its long history. A considerable portion of the May 4th youths were brought up in large extended families. In comparison, modern students were mostly brought up in urban nuclear families and were subject to much more liberal child-rearing practices. During the May 4th era, female students had to fight to wear modern hairstyles and to attend the same schools as males. In contrast, students in the 1980s could date rather freely. Moreover, many of the May 4th and the December 9th students were educated with a blend of traditional and modern curricula. In comparison, the students in the 1980s were educated with more Westernized curricula. Finally, modern students lived in the age of media technology. They were much more exposed, albeit superfi-

¹¹ One may simply view the traditionalism during the 1989 BSM as a nostalgic cultural revival. While people are nostalgic, this is not enough to explain the revival of a major cultural tradition. Major cultural revival must be accompanied by certain structural changes that either make that culture more relevant or facilitate the development of certain cultural institutions that fight for that culture. In a way, this article can be seen as an analysis of the structural conditions that revived some of the traditional cultural elements in China during the 1980s.

cially, to the Western lifestyle and mentality than earlier students. Therefore, we need other explanations.

THE THREE DIMENSIONS OF STATE-SOCIETY RELATIONS

I have shown that the 1989 BSM's rhetoric and activities were more traditional than those of the two earlier student movements. Before explaining this finding, I need to specify three structural dimensions of state-society relations that will be used to elaborate the comparison. These dimensions are the nature of the state, the nature of society, and the links or patterns of interaction between the state and society.

Scholars have proposed different ideal-types of states. For example, recently *Finer* (1997) has classified major regime types based on four institutional sources: palace, forum, church, and nobility. Their combinations yield six hybrid types. *Mann* (1988, 1993) identifies four types of regimes, differentiated by the level of a state's despotic power and infrastructural capacity to penetrate society. While modern democratic regimes are characterized by strong infrastructural but weak despotic power, modern totalitarian regimes have strong infrastructural and despotic power.

In their study of the democratic transition, *Linz and Stepan* (1996) identify five ideal-types of modern states: democratic, authoritarian, totalitarian, post-totalitarian, and sultanic. Different regime types have different paths and tasks in making the transition to democracy. Most relevant to this article are the authoritarian regime and the post-totalitarian regime. By authoritarian regime, *Linz and Stepan* refer to a political system that has no elaborated guiding ideology, but has a certain social and economic pluralism and a limited political pluralism, mainly a result of the state's weak capacity to penetrate society. A post-totalitarian regime is a decayed form of the totalitarian regime. In contrast to the authoritarian regime, this type of regime emphasizes rational decision making but is still officially committed to an elaborate ideology. It has a certain level of economic and social pluralism but a very limited political pluralism, due to the state's still-overwhelming presence. As it will be discussed in the next section, China was governed by authoritarian regimes during the May 4th and December 9th era, and by a post-totalitarian regime during the late 1980s. Because of its post-totalitarian nature, the Chinese state during the 1980s was more capable of carrying out repression than the state had been during the earlier student movements. The level of state unity—mainly elite unity—and the state's infrastructural capacity to penetrate society are thus the initial criteria of my comparison.

The second dimension of state-society relations is the structure or nature of society. Recent publications on the role of civil society in democratic transitions have greatly enhanced our understanding of this dimen-

sion (e.g., Ehrenberg 1999; Hall 1995; Keane 1988). The density and strength of intermediate organizations are crucial to the nature of society because intermediate organizations have many social movement–related functions. First, intermediate associations check state power. Without them, state power will be truly despotic, and people’s attention will be state centered (Tocqueville 1955). Second, intermediate organizations nurture bonds and contractual relations among social groups and create and sustain a shared identity and tradition. Poorly organized individuals tend to have little mutual understanding and little capacity to pursue their interests in a coordinated manner (Kornhauser 1959). While poorly organized movements tend to be driven by activities with strong emotional elements, stronger organizations make the rationalization of movement strategy possible. Finally, strong intermediate organizations facilitate sectional and segmental mobilization and prevent unitary national-level uprising. Organizations do play a crucial role in most microlevel participant recruiting processes, as resource mobilization theorists have argued (McCarthy and Zald 1973; McAdam 1982; Tilly 1978). However, different organizations tend to have distinctive interests and tend to nurture cross-cutting identities of their members. Thus, when the society is taken as a whole, the existence of strong intermediate organizations actually prevents unitary national-level mobilization. In the next section, I will show that, due to the state’s post-totalitarian nature, Chinese society during the 1980s had a very poor level of intermediate organizations compared to the eras of the May 4th and December 9th movements.

The third dimension of the comparison is where the state and society are linked. The importance of economic and political linkages has been conventional wisdom. The analysis in this article relies heavily on the ideational links between the state and society: on people’s perception of state legitimacy. State legitimacy has been defined in many ways (Habermas 1975; Finer 1997; Johnson 1982; Linz and Stepan 1996; Offe 1973; Wolfe 1977). There is also a significant difference between people’s perception of state legitimacy and the bases of state legitimation that state authorities use to justify their rule. Since I mainly deal with the population at large, not with state elites, this article focuses on people’s perception of state legitimacy.

Weber believes that tradition, charisma, and rationality are the three bases of human compliance (Bendix 1960, pp. 290–97; Weber 1978, p. 28). Weber’s classification is illuminating, but when it is used to analyze people’s perception of state power, it raises a problem. Not only in modern times, but even in the past, people’s perception of state legitimation involved a mix of emotions and rationality regardless of the ways in which a state tried to legitimize itself. For example, in historical China, the emperor was legitimized as “the son of the heaven,” yet dynasties were fre-

quently challenged and overthrown when heaven's son failed to provide basic services. The same is true when people support a charismatic modern state leader. They do so both out of affection and hoping for a reward. I therefore slightly modify Weber's classification system. In this article, sources of state legitimation are not classified in terms of basic types of human compliance, but in terms of how state power is justified: by a commonly accepted procedure, by the service that the state has provided, or by a future promise. Correspondingly, I define three types of state legitimation: legal-electoral, ideological, and performance legitimacy, respectively.

A state is likely to enjoy legal-electoral legitimacy when it takes laws as binding principles for all social groups, including state elites themselves, and when top leaders are popularly elected on a regular basis. Ideological legitimacy means that a state power is based on a grand vision that rests on future promises to which a government is committed.¹² Performance legitimacy means that a state's right to rule rests on its continuing economic development, its observance of certain moral and ritual practice, or its maintenance of national defense. Finally, when citizens tie their hopes to the ability and personality of one or a few state leaders, the state enjoys charismatic legitimacy. Charismatic legitimacy can be supplementary to any kind of state legitimacy, but it tends to be an extreme form of ideological legitimacy.¹³ These are all ideal-types. In reality, a state seldom bases its survival on a single source of legitimacy. However, in a particular country and at a particular time, one source of state legitimacy tends to dominate.¹⁴

Many classic studies suggest that people interpret an event by following preexisting "schemata of interpretation" in their minds. They tend to be upset when an action violates their expectations and to become excited when an action resonates with what preexisted in their minds (Collins 1981; Garfinkel 1967; Goffman 1967). In this article, I take people's preexisting perceptions of state legitimation as such "schemata of interpreta-

¹² People may find similarities between legal-electoral and ideological legitimacy since legal-electoral legitimacy has democracy at its ideological base. However, they are different in two aspects. Democracy only promises a procedure to select leaders, not a utopian future. More important, my definition of legal-electoral legitimacy emphasizes the procedural, not ideological, aspect. Although democracy is the base of legal-electoral legitimacy, over time it is the commonly accepted procedure of leadership selection, rather than the value system behind it, that legitimizes a state.

¹³ Defined this way, people's perception of state legitimacy can be assessed by their views on a state's ideological claim, their satisfaction with a state's major performance, and their impression of the top state leaders.

¹⁴ During Mao's era, the dominant source of state legitimation was ideology. However, the state also based its survival on Mao's charisma, as well as on some public goods that the state delivered to the previously lower classes in society.

tion" in society. In this sense, sources of state legitimation may determine to what kinds of rhetoric and activities the people will be more receptive, and therefore what kinds will dominate a social movement. In the next section, I will show that while nationalism was a major source of state legitimation during the May 4th and December 9th eras, the Chinese during the 1980s saw that the state's legitimacy lay primarily in moral and economic performance rooted in China's traditional state-society relations. This mode of legitimation also contributed to the traditionalism during the 1989 BSM.

STATE-SOCIETY RELATIONS IN THE THREE STUDENT MOVEMENTS AND ALONG THE THREE IDENTIFIED DIMENSIONS

In the last section, I defined the three major dimensions of state-society relations. Now, I will draw upon these dimensions in giving a brief historical account of the different state-society relations operating during the May 4th movement of 1919, the December 9th movement of 1935-36, and the 1989 BSM. I begin with the May 4th movement.

The May 4th movement started eight years after the 1911 Republican Revolution. When it began, China consisted of two rival states, with the north controlled by warlords and the south by a coalition of warlords and the Guomindang. The May 4th movement was aimed against the pro-Japanese force in the northern government. China at that time had two major political parties, the Guomindang and the Jinbudang (the Progressive Party). While they were otherwise in political competition, both parties urged the northern government not to sign the Versailles peace treaty. The Progressive Party was generally conservative and often a supporter of the warlord government in the northern parliament. During the May 4th movement, however, some members of the party, especially Liang Qichao, supported the students. Before the movement, the Guomindang had split into a southern revolutionary and a northern parliamentary faction, neither of which really initiated the movement. But individual Guomindang members, such as Cai Yuanpei, Wu Zhihui, Zhang Ji, Ye Chucang, Dai Jitao, and Shao Lizi were behind the movement from the very beginning. Moreover, more Guomindang members from both the northern parliamentary and southern revolutionary factions came to support the movement after it started. In fact, when the center of the movement shifted to Shanghai, the Guomindang practically led the movement (Liu 1990). But support for the movement was not limited to political parties. Some powerful warlords associated with the northern government, such as Wu Peifu and Zhang Jingrao, also publicly expressed their sympathy for students. Thus, the students were actually backed by a polit-

ically powerful and organizationally resourceful political coalition. Still, if the northern government during the May 4th era was no less authoritarian and repressive than the post-totalitarian communist government of the 1980s, its state elites were certainly more fragmented.

The northern government's weakness was reflected in the limited degree to which its authoritarianism penetrated Chinese society. Between 1912 and 1914, that government had published a series of laws to restrict people's rights to free speech, assembly, association, and the press. However, such laws could not extend to provinces and cities controlled by southern revolutionaries, or to foreign concessions, where publication and association were not regulated by Chinese law. Even within its own sphere of political influences, the northern government had proved unable to eliminate organizations not in its control. Therefore, commercial organizations, worker unions, and other nongovernmental organizations were in place long before the May 4th movement. Some of these organizations, such as the Chamber of Commerce of China in Beijing and Shanghai, the Citizens' Diplomatic Association, and the Chinese Industry Association, played important roles during the May 4th movement. After Yuan Shikai's failure to restore a monarchy and his subsequent death in 1916, the Beijing government's capacity for control deteriorated even further. Newspapers, magazines, and student organizations controlled by the new intellectuals mushroomed in Beijing. At Beijing University, for example, over 20 new associations were formed before the rise of the May 4th movement. The Association of Personal Ethics (*Jindehui*), with a membership of over 500 people (over 25% of the university population), had organized virtually all the new intellectuals in the university. Organizations such as the New Tide Society, the Citizens' Magazine Society, and the Mass Media Research Society provided the leadership for the movement (Zhang et al. 1979).

Immediately before the rise of the December 9th movement, by contrast, China seemed more tightly controlled by a repressive government. For by 1935, the Guomindang had unified most parts of China and pushed the CCP to abandon its southern bases and retreat to the Shanbei area. In the meantime, the Guomindang also tried to eliminate the CCP in the cities. Hundreds of thousands of communists and their supporters were captured and killed, and most underground communist organizations in the cities were destroyed. This "white terror" also extended to Chinese society at large. After 1927, procommunist worker unions, student societies, and left-wing organizations were banned, and other social organizations were placed under governmental control.¹⁵ In Beijing, for example,

¹⁵ For example, whereas before 1927 Shanghai had been a heartland of left-wing working-class politics, in the years between 1927 and 1937 left-wing worker's movements

the universities, with the exception of Yanjing and Qinghua,¹⁶ were politically very quiet before 1935.

However, although the Guomintang regime was very repressive, a highly divided governing elite on the top, and the existence of a professional Leninist communist party at the bottom, strongly affected the government's capacity to penetrate the society. When China was unified in the late 1920s, most warlords were not militarily eliminated, but politically co-opted. Thus many warlords still controlled their own bases and troops. Some of them, such as Feng Yuxiang and Zhang Xueliang, had great sympathy for the December 9th movement because of their nationalistic feelings and because of the fact that their northern power bases were being undermined by the Japanese.¹⁷ Furthermore, the Guomintang had a revolutionary tradition rooted in its history and in its ideological underpinnings. Chiang Kai-shek was able to repress the communists but could do very little to highly influential left-wing party members. During the December 9th movement, left-wing and liberal Guomintang leaders, led by Song Qingling and He Xiangning, the widows of Sun Yat-sen and Liao Zhongkai, also vehemently supported the students' actions. Therefore the governing elites' attitudes toward the movement were highly divided.¹⁸

Meanwhile, although the communists were driven to Shanbei, the Guomintang was unable to defeat them completely. Even before they arrived in Shanbei, the communists advocated a united front with the Guomintang against the Japanese invasion. At the time, most government troops stationed near Shanbei were part of Zhang Xueliang's army, which had retreated from Dongbei since 1931. Thus, the communists' proposal strongly resonated among Dongbei Army officers who were homesick and deeply humiliated by their peaceful retreat from their home province. For

were so repressed that worker unions in Shanghai became controlled by local gangsters (Perry 1993, chap. 5).

¹⁶ Yanjing was an American private university, and Qinghua was established and funded by war reparations that China paid to the United States. In both universities, the Guomintang's influence was relatively weak and free associations were more tolerated.

¹⁷ This was especially true of Zhang Xueliang, who had come under a lot of public pressure since he followed Chiang Kai-shek's orders and withdrew his troops from Dongbei in 1931 without offering any resistance to the Japanese invasion.

¹⁸ Here is an example of how students benefited from the elite's divisions. On February 23, 1936, 44 Dongbei University students were arrested in Beijing. Student leader Song Li went to Zhang Xueliang for help. Zhang wrote a letter to the chief of the military police in Beijing and requested that he release the students. He also asked Song Li to act as his secretary and to deliver the letter. When Song was in Beijing, he was recognized by local authorities as one of the students on the arrest list, yet nobody dared to arrest him. Instead, Song was treated well, and all of the 44 arrested students were released as Zhang requested (Song 1982). This kind of politics would have been unimaginable during the 1989 BSM.

a period, a truce existed between local Guomindang troops and the Red Army.

In Beijing, by the end of 1934, the CCP had been completely destroyed as an organization. However, in early 1935, the CCP started to reestablish its underground organs through a peripheral organization called the Chinese Association of Military Defense. The members of the association used Yanjing and Qinghua Universities as bases to expand their influence. By the time the December 9th movement broke out, the CCP had reestablished its Beijing branch and controlled radical student networks (Chen 1982; Yao 1987). Most of these students turned out to be crucial leaders in the movement and later revolutionary veterans. Thus, even though the CCP headquarters at Shanbei was not aware of it, the CCP local in Beijing played a role in facilitating the December 9th movement.¹⁹ Moreover, soon after the movement began, the CCP strengthened its presence in the movement. The CCP gradually gained control over the movement in several major cities, particularly Beijing and Tianjin. Thus the December 9th movement was increasingly supervised by the CCP.

One point that I need to bring out is that the principal base of state legitimation during both the May 4th and December 9th periods was nationalism. Since the Opium War in 1840, foreign aggression had become an increasingly urgent problem for China. In the wake of foreign aggression, nationalist consciousness rose first among intellectuals and then among other urban populations. China's defeat in the Sino-Japanese War in 1894 was a crucial catalyst for radicalism. Having a strong state capable of defending China became a dream of the Chinese (Meisner 1967; Pusey 1983; Yu 1993). It was in this general atmosphere that Liang Qichao, one of the foremost reformers during the late imperial and early republican era, proclaimed that: "even if a government system deprives the people of much or all of their freedom, it is a good system so long as it is founded on a spirit of meeting the requirement of national defense" (see Nathan 1986, p. 62). It was in the same atmosphere that China's political elites staged reforms and rebellions that eventually overthrew the Manchu government and established the Republic. Finally, this same radical atmosphere still prevailed when Chinese intellectuals and students staged the May 4th and December 9th movements.

¹⁹ Edgar Snow (1966) insists that the December 9th movement was largely spontaneous and that the early leaders of the movement were "mostly Christian or Christian-trained youths . . . not a Communist among them." Snow seems to have underestimated the communists' role in initiating the movement. In fact, many leaders of the movement had joined the CCP before the movement began. Some of them, such as Yu Qiwei (Huang Jing), Yao Keguang (Yao Yilin), Wang Rumei (Huang Hua), and Song Li, were also among the frequent guests of Snow's family. See Liao, Zhang, and Liu (1993) for the political careers of these December 9th movement leaders.

As a post-totalitarian regime, the Chinese government during the 1980s still claimed its legitimacy in ideological terms—in terms of the Four Cardinal Principles written in the preamble of the Chinese Constitution.²⁰ However, after the reforms of 1978, China's urban population knew more and more of the outside world. In comparison with the development of the West, and especially of the newly rising East Asian neighbors, urban Chinese found that their government had delivered to them far less than had been promised. Coupled with bitter memories of the disastrous Cultural Revolution and with the new international wave of democratization, the younger generation of intellectuals and students became increasingly critical of the regime's ideological stance (Zhao 1997). This does not mean that most people during the 1980s perceived the state as illegitimate, however. What happened in China was a shifting of public perception of the basis of state legitimation from leftist ideology back to performance—mainly moral and economic performance.²¹ Particularly relevant to this article was the moral dimension of state legitimation, which owes its legacy largely to the Confucian principle of “governing by goodness” (e.g., Creel 1953; Yang 1991). According to this principle, rulers must fulfill certain materialistic and ritualistic obligations to show their responsibility to and love of their subjects, who in exchange offer conformity and loyalty. After the ideological legitimation of the state declined, moral and ritual performance became a popular criterion by which the people would judge their government. Thus, during the 1980s, many novels became bestsellers in China not because of their quality, but because their topic was on upright officials (*qingguan*).²² Similarly, popular sayings such as “If an official is not able to stand up for his people, he should go home to sell sweet potatoes” (*Dangguan buwei minzuozhu, buyo huijia maihongshu*) became cur-

²⁰ These are adherence to socialism, adherence to the leadership of the party, adherence to Marxism-Leninism and Mao Zedong thought, and adherence to the dictatorship of the proletariat.

²¹ During my interviews, I asked informants the question: “What was your view on the Four Cardinal Principles before the 1989 student movement?” Among the 60 people who responded to the question, 21 (35%) were strictly against them, whereas only 9 (15%) accepted them. The rest (50%) accepted them with reservations. For those who had conditionally accepted the Four Cardinal Principles, most of them said so only because they thought that the state was still effective in certain functions, especially in leading economic development. This finding is also confirmed by other similar studies during the 1980s (Zhao 1995). Therefore, during the 1989 BSM, at least among intellectuals and students, the Chinese state enjoyed a certain degree of performance legitimacy, but not ideological legitimacy.

²² Ke Yunlu's *Xinxing* (A new star) was perhaps the most famous such novel. It shows how a young county mayor had acted as a traditional upright official for the benefit of the local people. When the novel was turned into a TV mini-series, each Saturday evening when the mini-series was broadcast, the major streets in Shanghai looked empty (Lull 1991, chap. 6).

rent.²³ The same reasoning also made official corruption an extremely sensitive issue in China and a major point of contention during the 1989 BSM.

Nevertheless, although a crisis of ideological legitimacy could be widely felt in urban China during the 1980s, no rival political organizations or parties existed then, as had been the case during the earlier student movements. Moreover, in the late-1980s most of the top-level government offices were still controlled by CCP veterans who had joined the party long before the communists took power. These veterans were aware of China's economic backwardness, among other problems. However, they also believed that the problems did not result from communism or authoritarianism, but rather from inexperience. They realized that reform was an absolute necessity, but they never intended to give up what millions of their comrades had died for.²⁴ They were confident because they had an armed force still controlled by veteran senior officers. Thus, the legitimization crisis actually had no catastrophic impact on the higher echelons of the leadership. Western analysts have highlighted the factional nature of Chinese government during the 1989 BSM. But the crucial fact is that no top state leaders really supported students during the movement. Their differences focused more on strategies to stem the movement (Zhao 1999). This contrasts sharply with what happened during the May 4th and the December 9th movements, when powerful governing elites openly supported students.

China in the 1980s was much freer than it was during Mao's era. This created an opening for the rise of the 1989 BSM. However, this openness should not be exaggerated. In the early years when the communists took over power, they wiped out the independent social organizations and replaced them with party-controlled worker, woman, youth, and vocational organizations. New forms of nonofficial organizations emerged during the 1980s as China's reform deepened (White 1993). However, the new associations were still at a very rudimentary stage of development and not comparable to the organizations of the republican era. Most of these new organizations also had a semiofficial status and were never intended to mobilize against the state (Wang, Zhe, and Sun 1993). Thus, during the 1989 BSM, all the movement organizations emerged with no prior history,

²³ This sentence was from a drama about an upright official in traditional China who risked his career to serve his subjects. The drama became very popular as soon as it came out, and the quoted sentence became a catchphrase for the Chinese urban population during the mid-1980s.

²⁴ For example, in a speech at the CCP Central Advisory Commission Standing Committee, Chen Yun (1990) said that "Everyone knows that the Chinese revolution went through decades of hard struggle and saw the sacrifice of more than twenty million people, and only then was the People's Republic of China founded. The victory has not come easily." Chen Yun's speech was highly representative of most top CCP leaders' attitudes during the 1989 BSM.

and student mobilization was facilitated more by the ecology of the campuses and of Tiananmen Square than by the strength of organizations (Zhao 1998).

STATE-SOCIETY RELATIONS AND PATTERNS OF MOVEMENT ACTIVITIES DURING THE 1989 MOVEMENT

In the last section, I argued that China during both the May 4th movement of 1919 and the December 9th movement of 1935–36 had a fragmented state, divided elites, strong opposition parties or intermediate associations, and a state legitimation primarily based on the state's capacity for national defense, while the post-totalitarian communist state during the 1980s had a relatively unitary governing elite that held much tighter control over society, despite the fact that a large proportion of the urban population judged it by its moral and economic performance, not by its ideology. Now, I will analyze how these differences shaped rhetoric and activities of the 1989 BSM.

Because China in the 1980s had a less fragmented state and weaker intermediate associations than had existed during the May 4th and the December 9th movements, the students during the 1989 BSM were unable to organize freely and raise demands directly. In this setting, the students sought to avoid head-on repression by twisting or hiding their real demands and goals behind legitimate forms of collective action. Many activities during the 1989 BSM can be straightforwardly explained as strategies that aimed to create a "safe space" to lower the possibility of repression (Evans and Boyte 1992; Gurr 1986; Lichbach 1995; Opp and Roehl 1990).

Let us recall the picket line example I gave earlier. In the past, the communist government had treated student activism relatively leniently but had severely sanctioned worker activism or any attempt to build an alliance between students and workers. Moreover, when a student demonstration was not picketed, the urban youths, some of them hooligans, had tended to mingle with demonstrators, making the demonstration less orderly and threatening to turn it into a riot. To avoid allowing the government to justify repression with such side effects, the students over time developed the picket line strategy to ensure that a demonstration went on in an orderly fashion.²⁵

The same logic also explains why funeral protests, which were only

²⁵ For example, riots broke out in both Changsha and Xi'an during and after student demonstrations on April 22. Massive looting occurred in both cities. Around 20 houses and 10 vehicles were burned during the Xi'an riot. In the April 26th *People's Daily* editorial, the government claimed that student demonstrations had facilitated these riots. Student leader Cheng Zhen (1990, p. 175) also has given exactly the same account of why they tried to keep other sections of the Beijing population from joining the demonstrations during the early stage of the movement.

marginally used during the May 4th and the December 9th movements, became so important for the 1989 BSM. Before the December 9th demonstration, student activists explicitly rejected Helen Snow's (a foreigner) suggestion that students start their demonstration with a mock funeral by carrying a "corpse of North China" through the streets of Beijing (Isreal 1966, pp. 115–16). Why was a tactic that was rejected by traditional students heavily used by modern students? In fact, a few student activists in Beijing had begun preparing for the 1989 movement long before Hu Yaobang's sudden death. They would have started this movement anyway, regardless of Hu's physical condition. Yet Hu's sudden death provided an excellent opportunity. At the time of his death, Hu Yaobang was still a politburo member of the CCP, and therefore a top statesman. Since to mourn a government leader in China is a legitimate action, one that the government itself would also conduct, to start collective action by centering on Hu Yaobang's funeral shielded the intention of movement activists and made immediate repression impossible.

The students' fear of repression also manifested itself in other activities. On April 26, the government published a *People's Daily* editorial that labeled the movement as antirevolutionary turmoil instigated by a small number of black hands. When students demonstrated on April 27 in defiance of the editorial, however, the slogans that students chanted the most were "Long live the communist party!" and "Support socialism!" (Wu 1989, p. 62). They also chanted slogans centered on issues of official corruption and high inflation. The slogans were not radical, even from the perspective of communist leaders. This led many observers to conclude that the students were loyal to the system and that their movement was in many ways similar to the remonstrance of China's traditional intellectuals before an emperor. I do not deny that China's traditional political culture assigns particular social roles to intellectuals and that modern Chinese students were influenced by these scripts. However, the rhetoric that students used during the April 27th demonstration makes for a somewhat misleading example. As early as April 18, student activists had unveiled seven demands at Tiananmen Square, at the center of which were calls for the freedom of speech, the press, and association. These demands guided the movement from that point onward. That the slogans students used during the April 27th demonstration were strikingly different from the earlier ones is best explained by considerations of strategy. By demonstrating on the street after the *People's Daily* editorial, the students challenged the regime's authority. However, by restricting their rhetoric to conformist slogans and to a few major social problems of common concern, the movement activists were able to mobilize more students from within and public sympathy from without, while denying the government an excuse for repression.

Thus far I have explained how the state's repressive and unitary nature pushed social movement actors during the 1989 BSM to use conformist strategies to expand the potential for mobilization and lessen the possibility of an immediate repression. However, strategy cannot be the only explanation. During the movement, many students bit their fingers to write slogans in blood; one of my informants reported that he had hit and smashed his own face in anger; a student leader recalled that he saw a woman with tears, snivel, and hair all over her face kneeling outside the student headquarters at Tiananmen Square crying, chanting, and begging the students not to leave Tiananmen Square during late May.²⁶ In fact, emotional displays constituted a major part of movement activities. If I were a social psychologist, I would certainly attribute the domination of cultural activities during the 1989 BSM to the emotional nature of the movement. The logic is straightforward: when people act emotionally, they tend to act out what they are most familiar with, to act according to those behavioral codes that have already been imprinted in their minds and to which they have become habituated (Kitayama and Markus 1994; Triandis 1989).

However, such an explanation could not constitute the whole story either. We know that a huge number of people participated in the 1989 BSM in an increasingly unorganized fashion (Chen 1996). Also, most movement activities were carried out in a few places, such as Tiananmen Square, the Triangle at Beijing University, or the Third Student Dining Hall at the People's University (Zhao 1998). During the movement, people frequently visited these places after work, after dinner, or anytime as they wanted. Most of them simply acted as sympathetic bystanders. When so moved, a few of them would stage actions of their own. The relatively unorganized and place-centered nature of these activities made the relationship between movement actors and bystanders very much like that between "street theater actors" at a fair and the people who visit that fair. Someone who had visited Tiananmen Square during the movement was likely to encounter several independent protesting activities, either simultaneously or sequentially, and a huge crowd of sympathetic bystanders mingling inside and looking for something interesting. Those who acted on the stage might follow traditional, modern, or foreign scripts of social movements. They might also act out of impulse, calculation, or a combination of both. For the most part, their forms of actions and motivations did not matter at this level because most of these activities came and went without having a

²⁶ In my interview with this student leader, he insisted that during the martial law period Tiananmen Square was like a mental hospital. He usually had a clear mind while outside the square. However, each time he returned to the square and saw many people acting this way he became emotional and confused.

substantial impact on the movement. Yet some activities, like the three students kneeling and the hunger strike, touched the hearts of hundreds of thousands of sympathetic observers and shaped the movement dynamics. I found that it was at this level that movement activities during the 1989 BSM appeared to be more traditional than those during the May 4th or the December 9th movements. Obviously, what mattered here was the psychology of those who were watching the protesting activities.

In their study of the "Zoot-Suit" riot, Turner and Surace (1956) found that a symbol would not evoke uniform group action and unrestricted hostile behavior unless its presence aroused uniformly and exclusively unfavorable feelings toward an object under attack. If we consider movement rhetoric and activities as symbols, and the sources of state legitimation as such an object, the logic they uncovered is certainly applicable to my case. I argue that people's acceptance of particular movement rhetoric or activities is shaped by their perception of state legitimation. During the 1980s, moral performance was a major source of state legitimation. Therefore, people were more receptive to morally (culturally) charged activities and got angry when the government reacted improperly to these activities. For example, on April 22, three students knelt on the steps of the Great Hall of the People, but Li Peng failed to come out to receive their petition. This immediately triggered a citywide class boycott. In fact, according to several of my informants, many other activities took place around the same time, but none of them was able to capture the hearts of students and shape the future of the movement. An informant recalled what happened when Li Peng did not come out to receive the three kneeling students:

I never trusted the communist party. I did not believe that they would really do anything good. I knew that nothing would be accomplished even if Li Peng came out to meet us, yet I still wished that Li Peng would come out to say a few words. But, he did not come out. We felt extremely sad and many of us burst out crying. . . . At one point, I even punched my own face several times, and blood gushed out of my face. Then I was pulled out by several friends. . . . On my way out, I saw that students were crying everywhere. I also saw that a journalist was crying like a baby and chanting: "This is a bastard government. This is a shit government. This is a useless government." I also heard some students shout: "Down with Li Peng!"

This informant was a very radical student leader, and yet he wished that Li Peng would come out and meet them. As Pye (1990, p. 168) has pointed out, coming out and receiving a petition in such a circumstance was a ritual that even the old magistrate had to fulfill. Contrary to most students' expectations, Li Peng did not come out. On that day, what my informant and the journalist acted and cried out was actually a general sentiment—a sentiment of total disappointment in the government. They

had this sentiment not because the government was not democratic (even though some of them were fighting for democratization), but because a leader of the government failed to perform a proper ritual, a ritual upon which the state was now basing its legitimacy. It was out of this sentiment that most Beijing students participated in a general class boycott the following day.

During the April 27th student demonstration, as an informant recalled, Beijing citizens on the sidewalks responded enthusiastically when students shouted slogans against official corruption and high inflation. However, slogans such as "Free Press," "Free Association," and "Democracy" were poorly received. Consequently, my informant saw a gradual drift of the slogans shouted during the demonstration away from the issue of freedom and democracy toward those of official corruption and inflation. The reason for the differentiated response from the people on the sidewalks was clear. By the late 1980s, the major sources of state legitimacy were moral and economic performance; that is, most students and Beijing residents did not really care about how the government was formed, but rather about whether the government was moral and whether it could bring order and economic prosperity. Slogans that were centered on official corruption and inflation attracted a wide audience because they resonated with most people's expectations of the government.

Many other examples can be given of how people responded differently to various movement activities. During the hunger strike mobilization between May 11 and 13, radical students tried to mobilize students to join the strike by using different claims. In the University of Political Science and Law, the major reason was "the government has not yet rehabilitated our movement, but Dongchang starts to spread rumor again."²⁷ This was not successful. By noon of May 13, only seven or eight students had volunteered for the hunger strike. Beijing Normal University's major claim was "in order to advance the democratization in China, we voluntarily participate in a hunger strike." This was not convincing either. Very few students signed up. At the more radical Beijing University, activists had made different speeches to try to mobilize students. However, by the evening of May 12, only about 40 students had signed up to be hunger strikers. At this point, student activist Chai Ling made an emotional speech to tell her reason for a hunger strike. In her speech, she did not capitalize on the democratic aspect of the movement. Rather, she emphasized that "my hunger strike is for the purpose of seeing just what the true face of the government is, to see whether it intends to suppress the movement or to ignore it, to see whether the people have a conscience or not, to see if

²⁷ He Dongchang was the minister of education in China. The slogan was a response to his rather harsh comments about the movement made several days earlier.

China still has a conscience or not, if it has hope or not.”²⁸ Student leader Shen Tong was among the majority who were initially against the idea of a hunger strike. He recalled his impression after Chai’s speech (Shen 1990, p. 237):

Chai Ling’s speech was so personal and direct, yet it touched me as no other speech had. I had been trying all day to come to terms with the idea of the hunger strike, and no one had said anything that satisfied me. Chai Ling made me understand why students would want to make such a sacrifice. We were asking the government for dialogue and recognition. Were these two things worth dying for? I didn’t think so. But Chai Ling’s speech made me and many others realize that the hunger strike was about much more than those two demands.

After Chai’s speech, the number of students signed up for the hunger strike at Beijing University immediately went from around 40 to nearly 300. Rhetoric centered on the democratic goals of the movement was not effective, but a call for the students to use their own lives to test the government’s morality received a very enthusiastic response.

The above analysis explains why the 1989 BSM focused on moral issues even though the goal of the movement was democracy. Since moral performance was the basis of state legitimacy, activities that morally discredited the government posed a more fundamental challenge to the state, and thus were able to mobilize more people. Therefore, during the hunger strike, almost all the people in Beijing, even those working in state institutions, eventually supported students. Most Beijing residents also joined the mass resistance of the army after martial law was imposed. Most of them, especially those from peripheral government institutions, did not stand up to fight for a democratic China, instead they protested a central government that kept silent while students were dying of hunger and a government that tried to use troops to repress students. In short, many people participated in the movement because they were angry; they were angry because the government’s actions violated the people’s common notion of how a good government should act.

Now I want to return to the issue of why traditional languages and activities were less manifest during the two earlier student movements. The difference seems to be attributable to two factors. First, early this century, foreign aggression was the most urgent problem in China, and how to stop it was the focus of Chinese political elites. As stated before, the principal foundation of state legitimacy then lay in the state’s capacity to resist foreign aggression rather than in its moral conduct and its ability to regulate economy. In other words, while the nationalist goals of both

²⁸ This excerpt is from both Shen (1990, p. 237) and Han and Hua (1990, p. 198).

the May 4th and December 9th movements had actually challenged the regime's legitimacy, simple moral issues were at those moments less able to move people if they were not directly related to nationalism.

Second, the level of organizations in society was different in the earlier periods. All three student movements were initiated without particular organizations behind them. Emotional displays were also widespread during the May 4th movement of 1919 and the December 9th movement of 1935–36. However, in the two earlier movements, organized forces moved in to explore the situation, whereas during the 1989 BSM no such political forces existed. The strategic activities of political organizations also freed the two earlier student movements from emotion-ridden rhetoric and activities. Let me discuss a case during the December 9th movement to illustrate the point.

The December 9th movement started as a direct response to the Japanese aggression. Yet, the CCP supported and gradually took control over the movement as it developed. The CCP leadership saw the movement as an opportunity to gain back their political influence in cities and as an occasion to establish a united front with the Guomindang government against Japanese aggression. Obviously, the CCP was not interested in radical actions at that point.

Here, the turning point was occasioned by Guo Qing's death. Guo Qing, a high school student, was beaten and later died in a jail. The students in Beijing became very angry and organized a memorial service at Beijing University in protest. Local authorities banned the meeting and deployed a huge number of police outside the university. The students were very emotional. When a student rushed onto the stage and suggested carrying a coffin and demonstrating outside the university, most students followed. However, the demonstrators met head-on with repression. Only six to seven hundred students eventually managed to demonstrate outside the campus, where about a hundred were wounded and 53 were arrested, including several major student leaders. The action was immediately criticized by Liu Shaoqi as ultra-leftist.²⁹ Liu then reorganized the CCP Beijing Committee and dismissed those whose ideas were not congenial to the policy of the CCP Central Committee. He wrote a letter to students and requested them not to continue to act so radically. He also demanded that they resume classes and respect their teachers and parents.³⁰ Moreover, Liu asked students to differentiate between the Nanjing government

²⁹ Liu was then in charge of the CCP Northern China Bureau. He was the president of P. R. China before the Cultural Revolution, and died during the Cultural Revolution after being purged from the party.

³⁰ The article was reprinted in 1987 with a title: "Close-Doorism and Adventurism" (Liu 1987).

and that of the local military leader Song Zheyuan. He promptly raised the slogans: "Support the Chief of the Committee Song to lead the anti-Japanese resistance!" and "Support the 29th Armed Forces to fight against the Japanese!" He also contacted hundreds of Guomindang elites to gain their sympathy. Liu's efforts were successful. From then on, Beijing students no longer acted radically. Local military leader Song Zheyuan relented and released the arrested students. The movement changed its course after the CCP Central Committee gained control.

This example shows that Chinese students were indeed very likely to follow culturally embedded forms of collective action, such as the funeral ritual, during the movement. Yet, it also suggests that the impact of culture was not as deterministic as some cultural theorists have claimed. The students during the December 9th movement initiated a funeral demonstration in anger, but once it was suppressed, no more radical actions followed. On the contrary, the funeral demonstration was labeled as ultra-leftist, and the movement took a different course. The movement was able to drastically change its orientation because the student organizations were controlled either by CCP members or by left-wing students closely associated with the communists.

DISCUSSION

The 1989 BSM was dominated by many traditional languages and activities. This phenomenon has been explained as a result of characteristics of traditional Chinese culture. But to explain cultural manifestations as culture is tautological. Moreover, the interpretation cannot explain why traditional Chinese culture was less manifest during the May 4th movement of 1919 and the December 9th movement of 1935–36. I argue that its domination in the 1989 BSM was realized at the microlevel through two not always separable processes: a strategic and an emotional process. Both of these processes, however, can be further explained by three structural factors centered on state-society relations. In comparison with the two earlier movements, China in the 1980s had a relatively unitary governing elite with much tighter control over society, a weaker level of independent civil organizations, and a mode of legitimation based on moral and economic performance. Students had to conceal their real demands and goals behind culturally congenial forms of collective action in order to avoid an immediate state repression. Moreover, because of the weakness of independent social organizations, the 1989 BSM was composed of many small groups of poorly organized activists and a large number of (generally unorganized) sympathetic audiences. As a result, competing activities frequently occurred in the same time and location, and the perceptions of people watching movement activities became key determinants of what

types of activities dominated. I argue that people's perceptions of state legitimation shape their view on the proper political behaviors of movement activists and the government. During the 1980s, the source of state legitimacy in China had changed from ideology to moral performance, reviving it to what it had been in traditional China. Therefore those who were watching movement activities during the 1989 BSM were more receptive to morally charged activities and became angry when the government reacted insensitively to these types of activities. This further facilitated the domination of traditional Chinese culture during the 1989 BSM.

In this article, I have identified three macrostructural factors. I have also analyzed how these structural elements shaped the strategic choices as well as the emotions of movement participants. I have thus linked social structures with human agents that have the capacity for rational calculation as well as emotions. "Micro-macro links" have been a fundamental issue in sociology (Alexander et al. 1987). Currently, a dominant trend in sociology is to link structures to rational human individuals. The political opportunity structure theory, or David Snow's type of "strategic framing" analysis in social movement studies,³¹ and the new institutionalism in organizational studies (Adams 1996; Brinton and Nee 1998; Czada, Heritier, Keman 1996; DiMaggio 1988; Granovetter 1985) attest to the potency of this tradition. If I had followed this tradition, I could have simply argued that students during the 1989 BSM adopted ostensibly traditional cultural forms mainly because the regime's unity and repressiveness made such strategies safer. In the following, I will explain why I do not follow this line of reasoning.

Empirically, most observers may agree that the 1989 BSM was very poorly organized (e.g., Shen 1990; Tsou 1991; Zhou 1993). In certain events, there were so many independent activities that I feel very uneasy about assuming that some actions were a result of pure strategic calculation without even being able to identify the exact actors. Moreover, although the cultural explanation of the 1989 BSM is tautological, the observation that culture and emotions figured very importantly during the 1989 BSM is actually right and should not be neglected.

It may be argued that contemporary scholars who follow various "bounded-rationality" approaches also recognize the role of cultural and emotional factors in social actions (Boudon 1987). In social movement

³¹ For the political opportunity structure theory, see, e.g., Costain (1992), Gamson and Meyer (1996), McAdam (1996), Jenkins and Perrow (1977), Kitschelt (1986), Kriesi (1996), Kriesi et al. (1995), McAdam (1982), Porta (1996), Rucht (1990, 1996), Tarrow (1996, 1998), and Tilly (1978). For recent criticisms of this theory, see Jasper and Goodwin (1999). For the strategic framing theory, see Evans (1997), Johnston (1995), Ryan (1991), Snow et al. (1986), Snow and Benford (1988, 1992), and Williams and Williams (1995). Zou and Benford (1995) also apply frame analysis to the 1989 BSM.

studies, for example, Gamson and his associates have from very early on emphasized the role of the "injustice frame" and of cultural resonance in social movements (Gamson 1992; Gamson, Fireman, and Rytina 1982). Snow and his colleagues also argue that frames promoted by activists must correspond to the values and experiences of larger groups to attract wider support. Tarrow (1998, pp. 111–12), in his second edition of *Power in Movement*, adds a section that discusses the role of emotions in social movements. However, their understanding of culture and emotions is different from cultural theorists. When rational choice scholars talk about injustice, moral commitment, ritual, loyalty, and behaviors with a strong emotional dimension in a particular culture, they view them functioning as "solidarity incentives" (Oberschall 1998), or as a "tool kit" (Laitin 1988; Swidler 1986; Tarrow 1994, p. 122; Zald 1996) that assists people's rational calculation. On the other hand, for scholars who study culture and emotions, it is an indisputable fact that cultural and emotional forces also operate at a subconscious level as habits and instincts (Barbalet 1998; Collins 1990; Goodwin 1997; Gordon 1990; Griffiths 1997; Jasper 1998; Kemper 1990; Scheff 1990).

My choice to link social structure with real humans with emotions as well as rationality reflects my quite different understanding of the problems of micro-macro links. I will restrict the discussion to a most relevant issue. Assuming the readers accept that rational creativity and emotions are essential to human behaviors and that individuals are not always able to rationally manage their emotions, then we can immediately identify three ways to link microlevel social actions to macrosocial structures. The first links macrostructures to a rational human model. In social movement studies, proponents of political opportunity structure theories adopt this strategy. The second connects structures to emotional humans. The recently emerging culture and emotion-centered studies on social movements and other contentious politics follow this strategy (Jasper 1997, 1998; Jasper and Poulsen 1995; Lindholm 1990; Scheff 1994). The last way links macrosocial structures with real humans who have emotions as well as the capacity for rational calculation. It is this type of link that I have adopted.

Each of the three research strategies has its own values. When we link macrostructures with rational (or emotional) humans, we are actually interested in a more formal question, that of the amount of variations that a particular theory can explain when one assumes humans to be rational (or emotional). When we connect macrostructures with real humans with emotions and rationality, we are interested in explaining empirical reality in a truthful manner. While the former is an analytical model, the latter is an empirical model. Let me provide a closely related example from other disciplines. Physics is almost completely dominated by analytical ap-

proaches, because in physics analytical models such as Newton's law and Maxwell equations can describe the reality so well that no other approaches are necessary. In population biology, however, the situation becomes more complicated. For example, the Malthus equation captures the exponential nature of population growth. Yet few populations can grow exponentially due to population-specific constraints (such as food, density dependent diseases, inter- and intrapopulation competition, predators, climate, etc.). Therefore, population biologists have developed two modeling strategies. While analytical models (most start with the Malthus model) aim to capture formal essences of population dynamics, empirical models (usually the extension of life table methods) incorporate as many empirical complexities as possible to describe population dynamics in the closest possible manner. The purpose of this article is not to develop a formal theory, but to understand empirically why the 1989 BSM adopted a particular pattern of rhetoric and symbolic activities. Therefore, I link empirically determined social structures with real humans who have both rationality and emotions. This type of linkage, however, is not my invention. Indeed, as Stinchcombe (1975, p. 27) has commented, the secret to the success of George Herbert Mead and Max Weber lies in "the fact that they make both people and structures real."

I argue that an empirically motivated structural analysis on social movements should not be linked to a rational human model, but to real humans. This does not mean that empirically motivated studies are not able to deal with the issue of rationality and emotions in social movements. This article shows that structural factors such as the nature of a state, the level and strength of intermediate organizations in society, and the sources of state legitimation all have motivational and microlevel behavioral consequences. The 1989 BSM was more driven by activities with strong emotional elements rooted in Chinese culture than the two earlier student movements under republican China because it developed under a post-totalitarian state with higher state unity and the capacity to penetrate society—a society with a very poor level of intermediate organizations and a state legitimation based on moral and economic performance. By pinning down these structural conditions, I explained why the 1989 BSM was more driven by culture and emotions than the two earlier student movements. Seen in this light, even the rise and domination of rational human models-based sociological theories in the West may reflect a set of structural transformations. That is, when the market dominates the economy, national politics center on regular elections, and well-defined legal codes penetrate almost every sphere of human life, the most important dimensions of human activities fit into accounting books. Consequently, rationality dominates social life, and Western sociological theories also increasingly rely on rationalist assumptions about humans. Yet,

what the traditionalism during the 1989 BSM teaches us is that cultural and emotional claims might gain new purchase, at least in the political sphere, under certain structural transformations in society.

REFERENCES

- Adams, Julia. 1996. "Principals and Agents, Colonialists and Company Men: The Decay of Colonial Control in the Dutch East Indies." *American Sociological Review* 61:12-28.
- Alexander, Jeffrey C., Bernhard Giesen, Richard Munch, and Neil J. Smelser. 1987. *The Micro-Macro Link*. Berkeley: University of California Press.
- Amsden, Alice. 1989. *Asia's Next Giant: South Korea and Late Industrialization*. New York: Oxford University Press.
- Baker, Keith Michael. 1990. *Inventing the French Revolution: Essays on French Political Culture in the Eighteenth Century*. Cambridge: Cambridge University Press.
- Barbalet, J. M. 1998. *Emotions, Social Theory, and Social Structure*. Cambridge: Cambridge University Press.
- Bendix, Reinhard. 1960. *Max Weber: An Intellectual Portrait*. Garden City, N.Y.: Anchor Books, Doubleday and Company.
- . 1968. *State and Society: A Reader in Comparative Political Sociology*. Boston: Little Brown.
- Blelloch, David. 1969. *State and Society in a Developing World*. London: C. A. Watts.
- Boudon, Raymond. 1987. "The Individualistic Tradition in Sociology." Pp. 45-70 in *The Micro-Macro Link*, edited by Jeffrey C. Alexander, Bernhard Giesen, Richard Munch, and Neil J. Smelser. Berkeley: University of California Press.
- Brinton, Mary C., and Victor Nee. 1998. *The New Institutionalism in Sociology*. New York: Russell Sage Foundation.
- Burns, John P. 1989. "China's Governance: Turbulent Environment." *China Quarterly* 119:481-518.
- Calhoun, Craig J. 1991. "The Problem of Identity in Collective Action." Pp. 51-75 in *Macro-Micro Linkages in Sociology*, edited by Joan Huber. Newbury Park, Calif.: Sage.
- . 1994. *Neither Gods Nor Emperors*. Berkeley: University of California Press.
- Casper, Gretchen, and Michelle M. Taylor. 1996. *Negotiating Democracy: Transitions from Authoritarian Rule*. Pittsburgh, Penn.: University of Pittsburgh Press.
- Chan, Anita, and Jonathan Unger. 1990. "China after Tiananmen." *Nation*, January 22.
- Chehabi, H. E., and Alfred Stepan. 1995. *Politics, Society, and Democracy*. Boulder, Colo.: Westview Press.
- Chen, Qiwu. 1982. "Guanyu Yaoerjiu Yundong de Yixie Qingkuang" [On the December 9th movement]. Pp. 60-171 in *Yaoerjiu Yundong Huiyilu* [A memoir of the December 9th movement]. Beijing: Renmin Chubanshe [People's Publishing House].
- Chen, Xiaoya. 1996. *The Democratic Movement on Tian-an-men Square, 1989*. Taipei: Fengyunshidai Publishing Company.
- Chen, Yun. 1990. "Speech to CPC Central Advisory Commission Standing Committee." Pp. 331-33 in *Beijing Spring, 1989: Confrontation and Conflict*, edited by Michel Oksenberg, Lawrence R. Sullivan, and Marc Lambert. Armonk: Sharpe.
- Cheng, Chu-yuan. 1990. *Behind the Tiananmen Massacre*. Boulder, Colo.: Westview Press.
- Cheng, Zhen. 1990. "Gongyun yu Xueyun, Minyun" [On the worker movement, student movement and democracy movement]. Pp. 174-84 in *Gongren Qilaile* [Workers are standing up]. Hong Kong: Hong Kong Workers' Union Education Publishing House.

- Chow, Tse-tung. 1967. *The May Fourth Movement: Intellectual Revolution in Modern China*. Stanford, Calif.: Stanford University Press.
- Collins, Randall. 1981. "On the Microfoundations of Macrosociology." *American Journal of Sociology* 86:984-1014.
- . 1990. "Stratification, Emotional Energy, and the Transient Emotions." Pp. 27-57 in *Research Agendas in the Sociology of Emotions*, edited by Theodore D. Kemper. Albany: State University of New York Press.
- Costain, Anne W. 1992. *Inviting Women's Rebellion: A Political Process Interpretation of the Women's Movement*. Baltimore: Johns Hopkins University Press.
- Creel, Herrlee Glessner. 1953. *Chinese Thought from Confucius to Mao Tse-tung*. Chicago: University of Chicago Press.
- Czada, Roland, Adrienne Heritier, and Hans Keman. 1996. *Institutions and Political Choice: On the Limits of Rationality*. Amsterdam: VU University Press.
- Derrida, Jacques. 1978. *Dissemination*, translated by Barbara Johnson. London: Athlone Press.
- . 1981. *Writing and Difference*, translated by Alan Bass. Chicago: University of Chicago Press.
- Diamond, Larry, Juan J. Linz, and Seymour Martin Lipset. 1988. *Democracy in Developing Countries: Latin America*. Boulder, Colo.: Lynne Rienner Publishers.
- DiMaggio, Paul J. 1988. "Interest and Agency in Institutional Theory." Pp. 3-21 in *Institutional Patterns and Organizations: Culture and Environment*, edited by Lynne G. Zucker. Cambridge, Mass.: Ballinger.
- Dittmer, Lowell. 1990. "Patterns of Elite Strife and Succession in Chinese Politics." *China Quarterly* 123:405-30.
- Ehrenberg, John. 1999. *Civil Society: The Critical History of an Idea*. New York: New York University Press.
- Esherick, Joseph W., and Jeffrey N. Wasserstrom. 1990. "Acting Out Democracy: Political Theatre in Modern China." *Journal of Asian Studies* 49:835-65.
- Evans, John H. 1997. "Multi-Organizational Fields and Social Movement Organization Frame Content: The Religious Pro-Choice Movement." *Sociological Inquiry* 67:451-69.
- Evans, Peter. 1995. *Embedded Autonomy: States and Industrial Transformation*. Princeton, N.J.: Princeton University Press.
- Evans, Sara M., and Harry C. Boyte. 1992. *Free Space: The Sources of Democratic Change in America*. Chicago: University of Chicago Press.
- Finer, Samuel E. 1997. *The History of Government from the Earliest Times*, vol. I. Oxford: Oxford University Press.
- Foran, John. 1997. *Theorizing Revolutions*. London: Routledge.
- Foucault, Michel. 1980. *The Order of Things: An Archaeology of the Human Sciences*. New York: Vintage.
- . 1994. *Power/Knowledge: Selected Interviews and Other Writings, 1972-1977*, edited by Colin Gordon. Brighton: Harvester Press.
- Furet, François. 1981. *Interpreting the French Revolution*. Cambridge: Cambridge University Press.
- Gamson, William A. 1992. *Talking Politics*. Cambridge: Cambridge University Press.
- Gamson, William A., Bruce Fireman, and Steven Rytina. 1982. *Encounters with Unjust Authority*. Homewood, Ill.: Dorsey Press.
- Gamson, William A., and David S. Meyer. 1996. "Framing Political Opportunity." Pp. 291-311 in *Comparative Perspectives on Social Movements*, edited by Doug McAdam, John D. McCarthy, and Mayer N. Zald. New York: Cambridge University Press.
- Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. Englewood Cliffs, N.J.: Prentice-Hall.

- Geertz, Clifford 1973. *The Interpretation of Culture: Selected Essays*. New York: Basic Books.
- . 1983. "Blurred Genres: The Refiguration of Social Thought." Pp. 19–35 in *Local Knowledge: Further Essays in Interpretative Anthropology*. New York: Basic Books.
- Goffman, Erving. 1967. *Interaction Ritual*. New York: Doubleday.
- Gold, Tom. 1986. *State and Society in the Taiwan Miracle*. Armonk, N.Y.: Sharpe.
- Goldstone, Jack A. 1991. *Revolution and Rebellion in the Early Modern World*. Berkeley: University of California Press.
- Goodwin, Jeff. 1989. "Colonialism and Revolution in Southeast Asia: A Comparative Analysis." Pp. 59–78 in *Revolution in the World-System*, edited by Terry Boswell. New York: Greenwood Press.
- . 1997. "The Libidinal Constitution of a High-Risk Social Movement: Affectual Ties and Solidarity in the Huk Rebellion." *American Sociological Review* 62:53–69.
- Goodwin, Jeff, and Theda Skocpol. 1989. "Explaining Revolutions in the Contemporary Third World." *Politics and Society* 17:489–509.
- Gordon, Steven L. 1990. "Social Structural Effects on Emotions." Pp. 145–79 in *Research Agendas in the Sociology of Emotions*, edited by Theodore D. Kemper. Albany, N.Y.: SUNY Press.
- Granovetter, Mark. 1985. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91:481–510.
- Griffiths, Paul E. 1997. *What Emotions Really Are: The Problems of Psychological Categories*. Chicago: University of Chicago Press.
- Gurr, Ted Robert. 1986. "Persisting Patterns of Repression and Rebellion: Foundations for a General Theory of Political Coercion." Pp. 149–68 in *Persistent Patterns and Emergent Structures in a Waning Century*, edited by Margaret P. Karns. New York: Praeger.
- Habermas, Jürgen. 1975. *Legitimation Crisis*. Boston: Beacon Press.
- Haggard, Stephen. 1990. *Pathways from the Periphery: The Politics of Growth in the Newly Industrializing Countries*. Ithaca, N.Y.: Cornell University Press.
- Hall, John A. 1995. *Civil Society: Theory, History, Comparison*. Cambridge: Polity Press.
- Han, Minzhu, and Shen Hua. 1990. *Cries for Democracy: Writings and Speeches from the 1989 Chinese Democracy Movement*. Princeton, N.J.: Princeton University Press.
- Han, Xuchang. 1986. *Yaoerjiu Yundong Shiyao* [A history of the December 9th movement]. Beijing: College of the CCP Central Committee.
- He, Zhou, and Zhu Jianhua. 1994. "The Voice of America and China." *Journalism Monographs*, no. 143. Austin, Tex.: Association for Education in Journalism.
- Huan, Guocang. 1989. "The Roots of the Political Crisis." *World Policy Journal* 6: 609–20.
- Hunt, Lynn. 1984. *Politics, Culture, and Class in the French Revolution*. Berkeley: University of California Press.
- Israel, John. 1966. *Student Nationalism in China: 1927–1937*. Stanford, Calif.: Stanford University Press.
- Jabri, Vivienne. 1996. *Discourses on Violence*. Manchester: Manchester University Press.
- Jasper, James M. 1997. *The Art of Moral Protest: Culture, Biography and Creativity in Social Movements*. Chicago: University of Chicago Press.
- . 1998. "The Emotions of Protest: Affective and Reactive Emotions in Social Movements." *Sociological Forum* 13:397–424.
- Jasper, James M., and Jeff Goodwin. 1999. "Caught in a Winding, Snarling Vine: The Structural Bias Political Process Theory." *Sociological Forum* 14:27–54.
- Jasper, James, and Jane D. Poulsen. 1995. "Recruiting Strangers and Friends: Moral Shocks and Social Networks in Animal Rights and Anti-Nuclear Protests." *Social Problems* 42:493–512.

- Jenkins, J. Craig, and Bert Klandermans. 1995. *The Politics of Social Protest*. Minneapolis: University of Minnesota Press.
- Jenkins, J. Craig, and Charles Perrow. 1977. "Insurgency of the Powerless: Farm Worker Movements (1946–1972)." *American Sociological Review* 42:249–68.
- Jin, Yucheng, ed. 1996. "Piaopo de Honghaiyang: Wode Dachuanlian" [A waving Red Sea: my involvement in the great chuanlian]. Taipei: China Times Publishing Company.
- Johnson, Chalmers A. 1962. *Peasant Nationalism and Communist Power: The Emergence of Revolutionary China, 1937–1945*. Stanford, Calif.: Stanford University Press.
- . 1982. *Revolutionary Change*, 2d ed. Stanford, Calif.: Stanford University Press.
- Johnston, Hank. 1995. "A Methodology for Frame Analysis: From Discourse to Cognitive Schemata." Pp. 217–46 in *Social Movements and Culture*, edited by Hank Johnston and Bert Klandermans. Minneapolis: University of Minnesota Press.
- Katznelson, Ira, and Aristide R. Zolberg, eds. 1986. *Working Class Formation*. Princeton, N.J.: Princeton University Press.
- Keane, John. 1988. *Democracy and Civil Society*. London: Verso.
- Kemper, Theodore D. 1990. *Social Structure and Testosterone: Explorations of the Socio-bio-social Chain*. New Brunswick, N.J.: Rutgers University Press.
- Kitayama, Shinobu, and Hazel Rose Markus. 1994. *Emotion and Culture*. Washington, D.C.: American Psychological Association.
- Kitschelt, Herbert. 1986. "Political Opportunity Structures and Political Protest: Anti-Nuclear Movements in Four Democracies." *British Journal of Political Science* 16:57–85.
- Kornhauser, William. 1959. *The Politics of Mass Society*. New York: Free Press.
- Kriesi, Hanspeter. 1996. "The Organizational Structure of New Social Movements in a Political Context." Pp. 152–84 in *Comparative Perspectives on Social Movements*, edited by Doug McAdam, John D. McCarthy, and Mayer N. Zald. Cambridge: Cambridge University Press.
- Kriesi, Hanspeter, Ruud Koopmans, Jan Willem Duyvendak, and Macro G. Giugni. 1995. *The Politics of New Social Movements in Western Europe: A Comparative Analysis*. Minneapolis: University of Minnesota Press.
- Kristof, Nicholas D. 1990. "How the Hardliners Won." Pp. 171–84 in *Tiananmen: China's Struggle for Democracy*, edited by Winston L. Y. Yang and Marsha L. Wagner. Baltimore: University of Maryland Press.
- Laitin, David D. 1988. "Political Culture and Political Preferences." *American Political Science Review* 82:589–97.
- Liao Gailong, Pinxing Zhang, and Yousheng Liu. 1993. *Xiandai Zhongguo Zhengjie Yaoren Zhuanlue Daquan* [The biographies of contemporary Chinese politicians]. Beijing: China Broadcasting and Television Publishing Company.
- Lichbach, Mark Irving. 1995. *The Rebel's Dilemma*. Ann Arbor: University of Michigan Press.
- Lin, Nan. 1992. *The Struggle for Tiananmen*. Westport, Conn.: Praeger.
- Lindholm, Charles. 1990. *Charisma*. Oxford: Blackwell.
- Linz, Juan J., and Alfred Stepan. 1996. *Problems of Democratic Transition and Consolidation: Southern Europe, South America, and Post-Communist Europe*. Baltimore: Johns Hopkins University Press.
- Lipset, Seymour Martin. 1983. "Radicalism or Reformism: The Sources of Working-Class Politics." *American Political Science Review* 77:1–18.
- Liu, Alan P. L. 1996. *Mass Politics in the People's Republic: State and Society in Contemporary China*. Boulder, Colo.: Westview Press.
- Liu, Shaoqi. 1987. "Guanmenzhuyi yu Maoxianzhuyi" [Close-doorism and adventurism]. Pp. 62–65 in *Yaerjiu Yundong* [December 9th movement], edited by Zhonggong Beijing Shiwei Dangshi Ziliao Zhengji Weiyuanhui [Party History Council of the CCP Beijing Committee]. Zhonggong Dangshi Ziliao Chubanshe [CCP Party History Publication House].

- Liu, Yongming. 1990. *Guomindangren yu Wusi Yundong* [Guomindang and the May 4th movement]. Beijing: Zhongguo Shehuikexue Chubanshe [China Social Science Publishing House].
- Lull, James. 1991. *China Turned On: Television, Reform and Resistance*. London: Routledge.
- MaCartney, Jane. 1990. "The Students: Heroes, Pawns or Power-Brokers?" Pp. 3–23 in *The Broken Mirror: China after Tiananmen*, edited by George Hicks. Essex: Longman.
- Mann, Michael. 1988. *State, Wars and Capitalism: Studies in Political Sociology*. Oxford: Basil Blackwell.
- . 1993. *The Sources of Social Power*. Vol. 2, *The Rise of Classes and Nation-States, 1760–1914*. Cambridge: Cambridge University Press.
- Marks, Gary Wolfe. 1989. *Unions in Politics: Britain, Germany, and the United States in the Nineteenth and Early Twentieth Centuries*. Princeton, N.J.: Princeton University Press.
- McAdam, Doug. 1982. *Political Process and the Development of Black Insurgency, 1930–1970*. Chicago: University of Chicago Press.
- . 1996. "Conceptual Origins, Current Problems, Future Directions." Pp. 23–40 in *Comparative Perspectives on Social Movements*, edited by Doug McAdam, John D. McCarthy, and Mayer N. Zald. New York: Cambridge University Press.
- McAdam, Doug, Sidney Tarrow, and Charles Tilly. 1997. "Toward an Integrated Perspective on Social Movements and Revolution." Pp. 142–73 in *Comparative Politics: Rationality, Culture, and Structure*, edited by Mark Irving Lichbach and Alan S. Zuckerman. Cambridge: Cambridge University Press.
- McCarthy, John D., and Mayer N. Zald. 1973. *The Trend of Social Movements in America. Professionalization and Resource Mobilization*. Morristown, N.J.: General Learning Corporation.
- McDaniel, Tim. 1988. *Autocracy, Capitalism and Revolution in Russia*. Berkeley: University of California Press.
- . 1991. *Autocracy, Modernization and Revolution in Russia and Iran*. Princeton, N.J.: Princeton University Press.
- Meisner, Maurice J. 1967. *Li Ta-chao and the Origins of Chinese Marxism*. Cambridge, Mass.: Harvard University Press.
- Migdal, Joel S. 1994. "The State in Society: An Approach to Struggles for Domination." Pp. 7–34 in *State Power and Social Forces*, edited by Joel S. Migdal, Atul Kohli, and Vivienne Shue. Cambridge: Cambridge University Press.
- Morley, Felix. 1949. *The Power in the People*. Toronto: D. Van. Nostrand Company.
- Nathan, Andrew. 1986. *Chinese Democracy*. Berkeley: University of California Press.
- . 1990. *China's Crisis: Dilemmas of Reform and Prospects for Democracy*. New York: Columbia University Press.
- Oberschall, Anthony. 1998. "A Review on James Jasper's *The Art of Moral Protest: Culture, Biography and Creativity in Social Movements*." *Contemporary Sociology* 28:92–93.
- Offe, Claus. 1973. "The Abolition of Market Control and the Problem of Legitimacy (I)." *Kapitalistate* 1:109–16.
- Opp, Karl-Dieter, and Wolfgang Roehl. 1990. "Repression, Micromobilization, and Political Protest." *Social Forces* 69:521–47.
- Perry, Elizabeth J. 1993. *Shanghai on Strike: The Politics of Chinese Labor*. Stanford, Calif.: Stanford University Press.
- Perry, Elizabeth J., and Ellen V. Fuller. 1991. "China's Long March to Democracy." *World Policy Journal* 8:663–85.
- Porta, Donatella Della. 1996. "Social Movements and the State: Thoughts on the Policing of Protest." Pp. 62–92 in *Comparative Perspectives on Social Movements*, edited by Doug McAdam, John D. McCarthy, and Mayer N. Zald. New York: Cambridge University Press.

- Pusey, James Reeve. 1983. *China and Charles Darwin*. Cambridge, Mass.: Harvard University Press.
- Pye, Lucian W. 1990. "The Escalation of Confrontation." Pp. 162–79 in *The Broken Mirror: China after Tiananmen*, edited by George Hicks. United Kingdom: Longman.
- Rosen, Stanley. 1991. "Political Education and Student Response: Some Background Factors behind the 1989 Beijing Demonstrations." Pp. 416–48 in *Chinese Education: Problems, Policies, and Prospects*, edited by Irving Epstein. New York: Garland Publishing.
- Rucht, Dieter. 1990. "Campaigns, Skirmishes, and Battles: Anti-Nuclear Movements in the USA, France, and West Germany." *Industrial Crisis Quarterly* 4:193–222.
- . 1996. "The Impact of National Contexts on Social Movement Structure: A Cross-movement and Cross-national Comparison." Pp. 185–204 in *Comparative Perspectives on Social Movements*, edited by Doug McAdam, John D. McCarthy, and Mayer N. Zald. Cambridge: Cambridge University Press.
- Ryan, Charlotte. 1991. *Prime Time Activism*. Boston: South End Press.
- Scheff, Thomas J. 1990. *Microsociology: Discourse, Emotion and Social Structure*. Chicago: University of Chicago Press.
- . 1994. *Bloody Revenge: Emotions, Nationalism, and War*. Boulder, Colo.: Westview Press.
- Schmitter, Philippe C. 1993. *Some Propositions about Civil Society and the Consolidation of Democracy*. Wien: Institut d für Höhere Studien.
- Sewell, William H., Jr. 1985. "Ideologies and Social Revolutions: Reflections on the French Case." *Journal of Modern History* 57:57–85.
- Shen, Tong. 1990. *Almost a Revolution*. New York: Harper Perennial.
- Shirk, Susan L. 1993. *The Political Logic of Economic Reform in China*. Berkeley: University of California Press.
- Shue, Vivienne. 1988. *The Reach of the State: Sketches of the Chinese Body Politics*. Stanford, Calif.: Stanford University Press.
- Skocpol, Theda. 1979. *States and Revolutions: A Comparative Analysis of France, Russia, and China*. Cambridge: Cambridge University Press.
- . 1985. "Bringing the State Back In: Strategies of Analysis in Current Research." Pp. 3–37 in *Bringing the State Back In*, edited by Peter Evans, Dietrich Rueschemeyer, and Theda Skocpol. Cambridge: Cambridge University Press.
- Snow, David A., and Robert D. Benford. 1988. "Ideology, Frame Resonance, and Participant Mobilization." *International Social Movement Research* 1:197–217.
- . 1992. "Master Frames and Cycles of Protest." Pp. 133–55 in *Frontiers in Social Movement Theory*, edited by Aldon D. Morris and Carol M. Mueller. New Haven, Conn.: Yale University Press.
- Snow, David A., E. Burke Rochford Jr., Steven K. Worden, and Robert D. Benford. 1986. "Frame Alignment Processes, Micromobilization, and Movement Participation." *American Sociological Review* 51:464–81.
- Snow, Edgar. 1966. "Comment on the December 9th Movement." *China Quarterly* 26:171–72.
- Song, Li. 1982. "Zhongguo Xuesheng Geming Yundong de Laichao." [The wave of the Chinese student revolutionary movement]. Pp. 1–57 in *Yaoerjiu Yundong Huiyilu* [A recollection of the December 9th movement], edited by Shuxian Yang, Jiming Tang, Keqing Xu, and Donglin Wang. Beijing: Renmin Chubanshe [People's Publishing House].
- Steinberg, Marc. W. 1994. "The Dialogue of Struggle." *Social Science History* 18: 505–41.
- Stepan, Alfred. 1978. *The State and Society: Peru in Comparative Perspective*. Princeton, N.J.: Princeton University Press.
- . 1989. *Democratizing Brazil*. New York: Oxford University Press.
- Stinchcombe, Arthur L. 1975. "Merton's Theory of Social Structure." Pp. 11–33 in

- The Idea of Social Structure*, edited by Lewis A. Coser. New York: Harcourt Brace Jovanovich.
- Strand, David. 1990. "Protest in Beijing: Civil Society and Public Sphere in China." *Problems of Communism* 34:1-19.
- Sullivan, Lawrence R. 1990. "The Emergence of Civil Society in China, Spring 1989." Pp. 126-44 in *The Chinese People's Movement, Perspectives on Spring 1989*, edited by Tony Saich. Armonk: Sharpe.
- Swidler, Ann. 1986. "Culture in Action: Symbols and Strategies." *American Sociological Review* 51:273-86.
- Tang, Wenfang, and William L. Parish. 2000. *Urban Life in Reform China*. New York: Cambridge University Press.
- Tarrow, Sidney. 1994. *Power in Movement*. New York: Cambridge University Press.
- . 1996. "States and Opportunities: The Political Structuring of Social Movements." Pp. 41-61 in *Comparative Perspectives on Social Movements*, edited by Doug McAdam, John D. McCarthy, and Mayer N. Zald. New York: Cambridge University Press.
- . 1998. *Power in Movement*, 2d ed. New York: Cambridge University Press.
- Terdiman, Richard. 1986. *Discourse/Counter-Discourse: The Theory and Practice of Symbolic Resistance in Nineteenth-Century France*. Ithaca, N.Y.: Cornell University Press.
- Tilly, Charles. 1978. *From Mobilization to Revolution*. New York: Random House.
- Tocqueville, Alexis de. 1955. *The Old Regime and the French Revolution*. New York: Anchor Books, Doubleday.
- Triandis, H. C. 1989. "The Self and Social Behavior in Differing Cultural Contexts." *Psychological Review* 96:506-20.
- Tsao, King K. 1991. "Civil Disobedience and Dissent against the Party-State: An Eyewitness Account of the 1989 Chinese Student Movement." Pp. 153-72 in *Culture and Politics in China*, edited by Peter Li, Steven Mark, and Marjorie H. Li. New Brunswick, N.J.: Transaction Publishers.
- Tsou, Tang. 1991. "The Tiananmen Tragedy: The State-Society Relationship, Choices, and Mechanisms in Historical Perspective." Pp. 265-327 in *Contemporary Chinese Politics in Historical Perspective*, edited by Brantly Womack. Cambridge: Cambridge University Press.
- Turner Ralph H., and Samuel J. Surace. 1956. "Zoot-Suiters and Mexicans: Symbols in Crowd Behavior." *American Journal of Sociology* 62:14-20.
- Wade, Robert. 1990. *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*. Princeton, N.J.: Princeton University Press.
- Wang, Ying, Xiaoye Zhe, and Bingyao Sun. 1993. *Shehui Zhongjanceng: Gaige yu Zhongguo de Shetuan Zuzhi* [The social middle layer: the reform and development of associations]. Beijing: Zhongguo Fazhan Chubanshe.
- Wasserstrom, Jeffrey N. 1991. *Student Protests in Twentieth-Century China: The View from Shanghai*. Stanford, Calif.: Stanford University Press.
- Wasserstrom, Jeffrey N., and Elizabeth J. Perry, eds. 1992. *Popular Protest and Political Culture in Modern China*. Boulder, Colo.: Westview Press.
- Weber, Max. 1978. *Weber: Selections in Translation*, edited by W. G. Runciman. Cambridge: Cambridge University Press.
- Weil, Frederick D., ed. 1996. *Research on Democracy and Society: Extremism, Protest, Social Movements, and Democracy*, vol. 3. Greenwich, Conn.: JAI Press.
- White, Gordon. 1993. "Prospects for Civil Society in China: A Case Study of Xiaoshan City." *Australian Journal of Chinese Affairs* 29:63-87.
- Wickham-Crowley, Timothy P. 1992. *Guerrillas and Revolution in Latin America: A Comparative Study of Insurgents and Regimes Since 1956*. Princeton, N.J.: Princeton University Press.
- Williams, Gwyneth I., and Rhys H. Williams. 1995. "'All We Want Is Equality':

- Rhetorical Framing in the Fathers' Rights Movement." Pp. 191–212 in *Images of Issues: Typifying Contemporary Social Problems*, 2d ed. Edited by Joel Best. New York: Aldine de Gruyter.
- Wolfe, Alan. 1977. *The Limits of Legitimacy*. New York: Free Press.
- Wong, R. Bin. 1997. *China Transformed: Historical Change and the Limits of European Experience*. Ithaca, N.Y.: Cornell University Press.
- Wu, Mouren. 1989. *Bajiu Zhongguo Minyun Jishi* [Daily reports on the movement for democracy in China].
- Xu, Deheng. 1987. *Weile Minzhu yu Kexue* [For science and democracy]. Beijing: Zhongguo Qingnian Chubanshe [China Youth Publishing House].
- Xuese de Limin [The bloody morning]. 1989. Hong Kong: Chi Keung Publishing.
- Yang, Chung-fang. 1991. "Conformity and Defiance on Tiananmen Square: A Social Psychological Perspective." Pp. 197–224 in *Cultural and Politics in China*, edited by Peter Li, Steven Mark, and Marjorie H. Li. New Brunswick, N.J.: Transaction Publishers.
- Yang, L. Dali. 1996. *Calamity and Reform in China*. Stanford, Calif.: Stanford University Press.
- Yao, Yilin. 1987. "Yaoerjiu Yundong de Huiyi" [A memoir on the December 9th movement]. Pp. 321–31 in *Yaoerjiu Yundong* [December 9th movement], edited by Zhonggong Beijing Shiwei Dangshi Ziliao Zhengji Weiyuanhui [The Party History Council of the CCP Beijing Committee]. Zhonggong Dangshi Ziliao Chubanshe [CCP Party History Publication House].
- Yeh, Wen-Hsin. 1996. *Provincial Passages: Culture, Space, and the Origins of Chinese Communism*. Berkeley: University of California Press.
- Yu, Ying-shih. 1993. "The Radicalization of China in the Twentieth Century." Pp. 125–50 in *China in Transformation*, edited by Tu Wei-ming. Cambridge, Mass.: Harvard University Press.
- Zald, Mayer N. 1996. "Culture, Ideology, and Strategic Framing." Pp. 261–74 in *Comparative Perspectives on Social Movements*, edited by Doug McAdam, John D. McCarthy, and Mayer N. Zald. New York: Cambridge University Press.
- Zhang Yunhou, Xuli Yin, Qingxiang Hong, and Yunkai Wang. 1979. *Wusi Shiqi de Shetuan* [Mass organizations during the May 4th era]. Beijing: Sanlian Publishing House.
- Zhao Dingxin. 1995. *Reform and Discontent: The Causes of the 1989 Chinese Student Movement*. Ph.D. dissertation. Department of Sociology, McGill University.
- . 1996. "Foreign Study as a Safety-Valve: The Experience of China's University Students Study Abroad in the Eighties." *Higher Education* 31:145–63.
- . 1997. "Decline of Political Control in Chinese Universities and the Rise of the 1989 Chinese Student Movement." *Sociological Perspectives* 40:159–82.
- . 1998. "Ecology of Social Movements: Student Mobilization during the 1989 Pro-democracy Movement in Beijing." *American Journal of Sociology* 103:1493–529.
- . 1999. "State Legitimacy, State Policy, and the Development of the 1989 Beijing Student Movement." *Asian Perspective* 23:245–84.
- Zhao, Dingxin, and John A. Hall. 1994. "State Power and Patterns of Late Development: Resolving the Crisis of the Sociology of Development." *Sociology* 28:211–29.
- Zheng, Yi. 1993. *Lishi de Yibufen—Yongyuan Jibuchu de Shiyifengxin* [Part of the history—eleven forever undeliverable letters]. Taipei: Wanxiang Publishing Company.
- Zhou, Kate Xiao. 1996. *How the Farmers Changed China: Power of the People*. Boulder, Colo.: Westview Press.
- Zhou, Xueguang. 1993. "Unorganized Interests and Collective Action in Communist China." *American Sociological Review* 58:54–73.
- Zou, Jiping, and Robert D. Benford. 1995. "Mobilization Processes and the 1989 Chinese Democracy Movement." *Sociological Quarterly* 36:131–56.

Embedded Altruism: Blood Collection Regimes and the European Union's Donor Population¹

Kieran Healy
Princeton University

Blood donation is often cited as a perfect example of altruism. But blood must be collected as well as donated, and the organizational basis of the blood supply has been largely neglected. This article is a comparative study of blood collection regimes in Europe. Regimes are found to affect donation rates and donor profiles. When the Red Cross collects blood, donation is tied to religious activity and other volunteering, unlike state and blood bank systems. This study argues that collection regimes produce their donor populations by providing differing opportunities for donations. The analysis contributes to an institutional perspective on altruism and highlights the need to attend to the socially embedded nature of altruistic as well as self-interested action.

INTRODUCTION

Human blood is scarce, valuable, and much in demand, yet it is supplied by voluntary donors who receive nothing for their trouble. Blood donation has therefore been seen as "perhaps the purest example" of altruistic behavior (Elster 1990, p. 46). Its symbolic resonance—an anonymous gift of life to an unknown recipient—only makes it more likely to be mentioned in the same breath as altruism or volunteering (see, e.g., Radin 1996, p. 96; Etzioni 1988, p. 75; Walzer 1983, p. 91). Those in charge of the blood supply routinely stress that very few people give blood, and so we tend to think of donors as special people. The surprisingly small

¹I am indebted to Bruce Western for his comments and advice. I also thank Paul DiMaggio, Bob Wuthnow, Viviana Zelizer, and my fellow participants in Princeton's Empirical Research Seminar for discussion and comments on earlier versions of this article. Three *AJS* reviewers made a number of very helpful suggestions. An earlier version of this article was presented at the 1998 meetings of the American Sociological Association in San Francisco. Direct correspondence to Kieran Healy, Department of Sociology, 2-N-2 Green Hall, Princeton University, Princeton, New Jersey 08544-1010. E-mail: kjhealy@princeton.edu

amount of research on blood donation shares this view and looks for the sources of a donor's "altruistic identity" (Piliavin and Callero 1991). For all these reasons, blood donors provide the perfect example to those interested in attacking the self-interested utility maximizer of neoclassical economics. *Homo economicus* would not give blood unless he was paid enough money; real-life donors do not reason in this way.

This article presents evidence for a different view. Economic sociologists argue that exchange must be understood by examining the social-structural framework that provides the incentives, opportunities, and constraints with which the actors think and work. Though we may morally prefer one to the other, this point should apply equally to both altruistic and selfish action. Granovetter's original argument about embeddedness criticized *both* over- and under-socialized pictures of the actor and argued for a focus on the social structures and relations that shape all action, rational and otherwise (Granovetter 1985, pp. 505–7). But the moral and rhetorical attractions of blood donation have insulated it from this kind of scrutiny. The institutional underpinnings of the blood supply have been almost entirely overshadowed by the image of the individual altruist. This article is a step toward redressing this imbalance. I argue that, when viewed comparatively, blood can be seen not so much as something that individuals donate, but as something that organizations collect.

All industrialized countries have a strong and permanent demand for blood. Large parts of their medical systems would very quickly collapse without it. Everything from emergency paramedical care to routine operations would become difficult or impossible. In addition, many people's lives depend on a constant supply of blood products. But different countries choose to meet this demand in different ways. Some manage to collect much more blood per capita than others, and they get it from different kinds of people. Some countries have a relatively small pool of regular donors and others a larger group of occasional suppliers. If the blood supply simply relied on the goodwill of individual altruists, it is not clear why such variation should exist. Yet there has been next to no empirical investigation of these cross-national differences.

I analyze a large survey that contains information on patterns of blood donation in the European Union. I describe and discuss blood collection practices within the European Union and identify three relevant sources of variation: (1) the organization in charge of collecting blood (namely, the state, the Red Cross, and blood banks), (2) the presence or absence of a volunteer donor group within a country, and (3) the presence or absence of for-profit plasma collection. I draw on research about blood donation and other kinds of volunteering to derive hypotheses about these organizational differences. The analysis shows that there are stable patterns of variation across different systems: different organizations collect their

blood from different kinds of people. Moreover, the act of blood donation looks very different under different systems. In Red Cross regimes, blood donation is tied to participation in religious organizations and resembles other kinds of volunteering. This is not true of countries where blood is collected by independent blood banks or the state. These findings contribute to a comparative, institutional approach to giving and volunteering (Salamon and Anheier 1998; Wuthnow 1991*b*; DiMaggio and Anheier 1990) and suggest that we should treat altruism with the same sociological skepticism that has long been applied to self-interest.

THE ELUSIVE ALTRUIST: STUDIES OF BLOOD DONORS

The need for a comparative, institutional perspective on blood donation, and altruism in general, can be seen from the findings of the existing literature on individual donors. Studies generally try to establish the demographic characteristics and motivations of donors. When asked, most donors will give some altruistic reason for giving, often citing feelings of community attachment or some commitment to the common good as their motive. Researchers have tried to correlate these motives with the demographic characteristics of the donor. Reviews of research on donors find that the typical donor is a white male in his thirties with above-average income and educational attainment (Oswalt 1977), though the gender gap in donation may not be as wide today as it was in the past (Piliavin 1990).

These findings tend to mislead us about the special character of donors. They "suggest that there is a class of people that makes donations and another class that does not" (Roberts and Wolkoff 1988, p. 170). The reality is that, more often, "many people with the characteristics of the typical donor are unlikely to be donors. Many with the characteristics of the typical nondonor do in fact donate" (Roberts and Wolkoff 1988, p. 170). This should make us suspicious of simply thinking that the key to understanding altruism is to be found in the personal characteristics of donors.

There are some strictly individual constraints on donation. Women give blood less often than men for medical reasons. Women are lighter than men on average. They are also more prone to anemia and can become pregnant. Each of these conditions disqualifies one from donating, and so there are fewer women in the pool of potential donors. Older people are also more likely to be excluded from the donor pool for medical reasons. This does not explain why better-off and better-educated people give more. Nor does it explain why relatively few eligible people give in the first place. The motives of donors are clearly important but should not stop us from asking how the institutional setting—the organization of recruitment, collection, and publicity—might make it more or less difficult for some kinds of people to donate blood.

Studies of donor motivation do sometimes recognize the role of institutions, though their research design usually prevents them from investigating this role properly. Asking how more blood might be collected, Roberts and Wolkoff (1988) recommend that the structure of incentives offered to donors be changed, as opposed to searching ever harder for elusive altruists. Piliavin and Callero (1991) follow first-time donors longitudinally and report on donors with different degrees of experience. They develop an analysis of how a person grows into a "donor-role." But they also recognize that other, nonindividual factors are important. They give evidence that both personal networks and simple organizational differences have important effects on donation rates. If many of your friends are donors, you are likely to be a donor as well. The accessibility of blood centers—whether collection points are mobile or fixed, for example—also affects whether people give. However, their research design confines them to the United States, and so the effect of large-scale institutional variation is outside the scope of their study.

Hypotheses about Individual Donors

The findings from this literature are easy to summarize. The data analyzed here allow us to replicate previous work in this area in a cross-national research design. Studies have found a reliable "modal profile" for blood donors and a similarly typical pattern of altruistic motives (Piliavin 1990; Oswalt 1977; Oswalt and Hoff 1975; London and Hemphill 1965). The link between the two is not so clear, however, as many who donate do not fit the profile. On the basis of these studies, the expected demographic characteristics of individual donors can be summed up in the following hypotheses:

HYPOTHESIS 1.—*The modal donor is a male in his thirties.*

HYPOTHESIS 2.—*The odds of donating blood increase with educational attainment.*

HYPOTHESIS 3.—*The odds of donating blood increase with income.*

Piliavin and Callero's (1991) work suggests that people are more likely to donate blood if they know other donors, or if they know people who have received transfusions (or other blood products). Similarly, Drake, Finkelstein, and Sapolsky (1982, pp. 81–83) report that those who are "close to blood needs" will be more likely to donate. We should expect typical network effects here: if all your friends are blood donors, you are likely to be one too. If you know a hemophiliac, you should also be more likely to have given blood at some point.

HYPOTHESIS 4.—*The odds of donating blood increase if you know anyone (including oneself) who has received a blood transfusion.*

I already noted that strictly medical reasons do not explain why better-

educated and better-off people are more likely to give. Arguing that such people are more likely to be altruistic does not seem to take us much further. Individual motives for giving blood may be less important than institutionalized methods of collecting it. I now examine this possibility in detail.

BLOOD COLLECTION REGIMES IN THE EUROPEAN UNION

Richard Titmuss's *The Gift Relationship* (1971, 1997) remains the only study of cross-national institutional variation in the blood supply. It is still frequently cited for its finding that voluntary donation is both more socially just and economically efficient than for-profit exchange in blood. But the book is almost 30 years old and out of date in many important respects. Titmuss compared the American and British systems as they existed around 1969, when it was still possible to sell your blood in the United States. There has been essentially no commercial collection of whole blood in the United States since 1974, a policy change brought about in large part by the book itself. Though still a voluntary system, institutional conditions in the United Kingdom have also changed substantially (LeGrand 1997). And the events surrounding the emergence of HIV in the blood supply seriously challenged many of Titmuss's central claims. In particular, the AIDS disaster showed that the relationships blood collection organizations had with their suppliers and recipients were subtler than Titmuss realized. Relying on voluntary donors did not protect against HIV contamination, and in some ways, the commercial sector reacted more responsibly to the crisis than the voluntary sector (Healy 1999).

Although the details of his argument no longer apply, Titmuss rightly pointed away from individuals and toward the system as a whole. He did not think that Britain's blood supply was maintained by a nation of saintly individuals. Rather, their altruism was socially sustained through the structure of the health system.

How is the collection of blood organized in Europe today? In spite of a common EU policy encouraging voluntary donation, we find that rates of donation and modes of organization differ considerably across countries.

Cross-National Variation in Giving

Table 1 shows the percentage of people in each country who have ever given blood, as reported in the 1993 Eurobarometer survey (Reif and Marlier 1994). The rate ranges from 14% in Luxembourg to 44% next door in France. This wide variation in donation rates is interesting. Why should

TABLE 1
RESPONDENTS WHO HAVE EVER GIVEN BLOOD,
BY COUNTRY (%)

Country	Donors	N
France	44	1,027
Greece	38	1,009
Denmark	34	1,004
United Kingdom*	32	1,064
Germany	30	2,110
The Netherlands	28	1,014
Ireland	27	1,066
Spain	24	986
Italy	21	1,052
Belgium	20	1,077
Norway	16	996
Portugal	16	1,000
Luxembourg	14	621

* Excluding Northern Ireland.

there be as much as 20 or 30 percentage points difference between France and Greece, on the one hand, and Luxembourg and Portugal on the other? If we think of donation as purely a question of individual motivation, it seems unlikely that some general propensity to generosity should vary quite so sharply across Europe. We should also be wary of writing the difference off to cultural variation, particularly given that countries that we might expect to fall together culturally (e.g., France and Luxembourg, Denmark and Norway) have dissimilar donation rates.

I argue that individual-level explanations cannot account for this variation. We should instead look to organizational variation to explain the differences.

Cross-National Variation in Organization

There are three relevant institutional features of the blood supply that vary across Europe: the *collection regime*, the presence of *volunteer donor organizations*, and the presence of *for-profit plasma collection*.

Three collection regimes.—In 1989, a European Community directive committed the European Union to securing its supply of whole blood and plasma from voluntary, unpaid donors. However, this commitment to voluntarism said nothing about the kind of organization that should do the collecting. At the same time, the Council of Europe commissioned a series of research papers on blood suppliers that were later published as white

TABLE 2

EUROPEAN BLOOD COLLECTION REGIMES

System Type	Country
State run	Britain
	France
	Ireland
Blood banks	Denmark
	Greece
	Italy
	Norway
	Portugal
	Spain
Red Cross	Belgium
	Luxembourg
	The Netherlands
	Germany

papers (Hagen 1993; van Aken 1993; Genetet 1998). One of these (Hagen 1993) is based on interviews with those responsible for the blood and plasma supply in each EU member state and describes the different sorts of organization that exist within these countries.

There are three blood collection regimes in Europe (Hagen 1993, p. 34–85).² First, in some countries, a national health service or nationally run blood organization collects all of that country's blood. Countries with this sort of system have no other collection agencies. Second, the Red Cross may have a monopoly on collection or control a majority of it (with a minority held by hospital or community blood banks). Third, blood banks may have a monopoly on collection or control a majority of it (with a minority held by the Red Cross). According to the Council of Europe, countries fall into these categories as shown in table 2.

The British, Irish, and French national health systems have a monopoly on blood collection in their respective countries. These countries make up the "state" category. In the "blood bank" category, Denmark is the only country where banks have a monopoly. The other members of this group (Greece, Italy, Norway, Portugal, and Spain) have blood bank majorities, with a Red Cross minority share. In Belgium, Luxembourg, and The Netherlands, the Red Cross has a monopoly on the supply. In Germany, the Red Cross has a majority share of the supply, and blood banks have

² Unless otherwise cited, information in this section comes from this Council of Europe white paper.

a minority. While precise data on market shares would have been ideal, information was available only about the relative predominance of the Red Cross and blood banks in those countries where they share responsibility for the supply.

Germany is the only such "mixed" system where the Red Cross holds a majority. It is an unusual case in other respects also. First, as of 1989, it was the only EU country that obtained some of its blood from paid suppliers. Second, there is a mixed system of paid and unpaid donation within the nonprofit sector. Most hospital and community blood banks pay between DM 30 and DM 50 (about \$18 to \$30) per donation. (Government policy limits payment to a DM 50 maximum.) In some areas, Red Cross collection centers also compensate their donors. Third, regional governments (the various *Länder*) have different collection policies. There are also significant differences between Eastern and Western *länder* (Hagen 1993, pp. 74–75).

Volunteer donor organizations.—Donors are organized in one or more national groups in Denmark, France, Greece, Italy, and Spain. For example, Denmark's organization "was founded in 1932 when a boy-scout movement established a corps of young adult boy-scouts who on a voluntary non-remunerated basis were willing to be called to hospitals to donate blood" (Hagen 1993, p. 58). The Voluntary Blood Donors of Denmark supplies hospitals with donors. In return, hospitals pay the local organization a small fee per donation, which is used for publicity and further recruitment efforts.

Italy has a slightly different form of organization. It resembles a community blood bank model rather than a hospital blood bank. The difference is that instead of sending donors to hospitals, the organization collects the blood itself and sends it to hospitals. The majority of blood banks in Italy are run by one of three donor organizations. The largest is called Associazione Volontari Italiani del Sangue (AVIS) and claims about 800,000 members. The others are the Fratres and Federazione Italiana della Associazioni Donatori di Sangue (FIDAS), which also claim large memberships. They are organized in different parts of the country and do not compete with one another.³

By contrast, although France and Spain have a national and a bank-based system respectively, important donor organizations also exist in these countries. In my analysis, therefore, I distinguish between a blood-banking system and the presence of a voluntary donor group, even though in some countries the donor groups may have a hand in running the blood banks.

³ Interestingly, this apparently strong voluntary activity contradicts the charge that Italy has no voluntary sector worth speaking of. See, e.g., Perlmutter (1991).

For-profit plasma collection.—Much of the debate about blood has contrasted the dangers of commercialism with the virtues of volunteering. While all blood suppliers sell blood products to hospitals or to one another, there is essentially no individual market for blood of the kind that existed in the United States in the 1960s. The commercial collection and processing of plasma is the main exception to this rule. Plasma can be extracted from whole blood or obtained separately through plasmapheresis.⁴

Plasma and whole blood collection overlap in complicated ways. In general, although nonprofit blood suppliers may collect plasma from voluntary donors, they do not process it any further. It is either used directly (as with whole blood) or sold to commercial plasma fractionators. Practices vary. Denmark, for example, has a state-run fractionation plant (although its capacity is small compared to the other, commercial plants in the country). The only exceptions of interest here are those countries where a company buys plasma directly from individual suppliers (rather than in bulk from regular blood suppliers who obtain it from donors). Spain and Germany both have such a system.⁵

BLOOD DONATION: ORGANIZATIONS AND VOLUNTEERING

Blood Donation as Voluntary Work

With this information about the structure of the EU's blood supply, we are in a position to work out some hypotheses about the effects of organizations on blood collection rates. We can think of blood donation as a special kind of volunteering that involves more than just money or time.⁶ Indeed, the physical and symbolic nature of the gift of blood is what makes it attractive as the perfect example of altruistic giving. How similar blood donation really is to regular volunteering is partly an empirical question (which I examine below), but it is close enough to suggest that what we know about other kinds of volunteering might also apply to blood.

If blood donation falls under the rubric of volunteering in general, then a country's donation rate should line up with its rate of volunteering. Greeley (1997) discusses the best available data for Europe, which come from the 1991 wave of the European Values Survey (EVS). We can test

⁴ Plasmapheresis takes longer than donating a unit of blood. It can also be done much more often: the body replaces lost plasma much faster than it does lost blood.

⁵ The world market for plasma is dominated by the United States. A number of U.S. companies buy plasma from individuals, serving both domestic and export markets. The potential impact of a very large commercial plasma sector on donation rates is considered below, when I discuss the relationship of blood donation to volunteering.

⁶ Lee, Piliavin, and Call (1999) find a weak positive correlation between giving blood and giving money or time.

the strength of the analogy between blood donation and other kinds of volunteering by comparing data from the two surveys.

HYPOTHESIS 5.—*Countries with high rates of voluntary activity in general will also have high rates of blood donation.*

Greeley also notes that “church attendance and membership in religious organizations correlate with volunteering [in Europe]. . . . Even in countries where religious activity is not high, as in the Scandinavian countries, religious behavior still has a significant impact on voluntary service” (Greeley 1997, p. 71). His data, together with other research (Wuthnow 1991a; Chaves 1998), suggest that involvement in organized religious activity encourages both religious and secular forms of volunteering. The same should be true for blood donation. I use the survey’s measure of church attendance to test this hypothesis.

HYPOTHESIS 6.—*People involved in religious organizations will be more likely to donate blood.*

Blood Collection and Organizational Scope

In general, blood is collected by public or nonprofit organizations that solicit voluntary donations from individuals. Why might some forms of organization be better than others at doing this? Research on how organizations solicit money shows that rates of giving tend to increase with age. Educational attainment also usually has a positive effect (Clotfelter 1993; Hodgkinson and Weitzman 1992; Jencks 1987). But the most consistent finding in this literature is organizational rather than individual: being asked to contribute is one of the most important determinants of giving in general (Hodgkinson and Weitzman 1992; Piliavin and Charng 1990). And although there are better and worse ways to ask, simply giving someone the opportunity to volunteer can be decisive. A survey of American blood donors carried out in 1975 confirms this general finding. As in other studies, when asked why they gave blood, donors tended to give altruistic reasons. But when nondonors were asked why they had never donated, the two most common responses were “I was never asked” and “There was no convenient opportunity” (Drake et al. 1982, pp. 76–113). Poor organization—rather than selfish motivation—kept people from giving.

We should expect some collection regimes to do better than others depending on the opportunities for giving that they offer. Larger, better-funded organizations will be in a position to publicize their needs more than smaller, more disaggregated ones. The publicity efforts of large organizations also benefit from economies of scale and from name recognition and the trust they engender. (The Red Cross symbol is instantly recognized the world over.) Large organizations may also find it easier to coor-

dinate effective recruitment drives. Costs can be redistributed from easy-to-recruit to costly-to-recruit regions. For these reasons, in our case, we should expect the Red Cross to be better than blood banks at attracting donors.

If simply reaching potential donors is important, then countries where the state has a monopoly on collection should do best of all. Other things being equal, a national system is more likely to be better funded, have a wider coverage, and have more recruitment options open to it than other kinds of organization. Operating within a national health system, the blood-collection organization should find it easier to integrate its activities into the general package of benefits provided by the state. Giving blood might more easily be seen as part of a general *quid pro quo*, part of the individual obligation incurred by the public goods provided by the state. State systems should also have an easier time getting access to other state organizations where donors might be found (such as universities, the civil service, and so on). The resources to run large-scale recruitment efforts may also be more easily available to state-run collection agencies.

HYPOTHESIS 7.—National health systems will attract the most donors, followed by Red Cross systems.

Blood banking systems are by nature more disaggregated than either state or Red Cross systems. The collection regime is made up of a number of blood banks, often tied to local hospitals, usually self-administered, and always serving some local population. This does not imply that blood banks will do worse than larger alternatives, although they should show more variation. Some local banks will be better run than others. In addition, blood banks offer a wider range of incentives to donors than either Red Cross or state systems. Some use an insurance system, where donors build up credit for their own operations by donating regularly themselves. Banking systems are also more likely to offer autologous donations, where patients build up a stock of their own blood solely for personal use. This variability in organizational style and incentive systems should be reflected in collection rates.

HYPOTHESIS 8.—Blood banking systems will show the widest range of variation in donation rates.

What about effects of commercialism? Titmuss argued that paid suppliers would drive out volunteers. Given that the blood supply today is essentially voluntary, it is difficult to test this claim. The best we can do is to examine whether countries with a commercial plasma sector differ from those without one. Because the plasma supply may be commercialized, but whole blood collection almost never is, the opportunity to voluntarily donate always exists. By Titmuss's logic, we should therefore expect some specialization by income.

HYPOTHESIS 9.—*In countries with a commercial plasma supply, the poor will be more likely to sell their plasma than donate their blood.*

Conversely, five countries in the sample (Denmark, France, Greece, Italy, and Spain) have donor organizations that are committed to the ideal of altruism. We saw that one of the main findings in studies of individual donors is that, while there is a "modal donor," many who fit the profile do not donate. I suggest that donor organizations will find it easiest to recruit from this group of nondonors. Thus, if the modal donor is a well-educated male, a donor group will find it easiest to recruit the many well-educated males who would not otherwise donate. These groups will skew the donor population toward the modal profile, rather than increasing the odds of donation for atypical individuals.

HYPOTHESIS 10.—*Where a donor organization exists, people already likely to donate blood will be even more likely to do so, but those already unlikely to donate will be less likely to do so.*

DATA AND METHODS

The Survey

I analyze the Eurobarometer survey carried out in the European Union, along with Norway and Finland, in 1994 (Reif and Marlier 1996). The survey contains data on blood donation across Europe. Respondents were asked a battery of questions about blood and plasma donation. People were asked for their opinion about the way blood and plasma are collected and handled, their reasons for donating and not donating, their understanding of the differences between blood and plasma, and their attitudes about buying and selling blood.

The survey sampled persons ages 15 years and over residing in the 12 member states of the European Union, as well as Norway and Finland. It was carried out through multistage national probability samples and national stratified quota samples during March through June of 1994. The complete data set contains 540 variables and 19,477 cases. In all, there are 31 questions about blood donation and related issues. Finland is not included in the analysis because all blood-related data for this country were missing.

Models and Variables

I first present the results of a series of logistic regression models for each country. In each case, the dependent variable measures whether the respondent has given blood in the last year (coded "1" for yes, "0" for no; overall, about 7% of the respondents had done so). I also report results for

a mixed-effects model of the pooled data, where the dependent variable is whether the respondent has ever given blood.⁷

The independent variables in the model represent characteristics of individual donors and institutional environments. The former are recoded from the survey. The latter were obtained from the interview data detailed in Hagen's (1993) Council of Europe white paper. Scores for institutional variables were attributed to each observation using the country codes in the data set.

The institutional setting is represented by three variables: First is the kind of collection system a country operates. Countries with state-run systems make up the omitted category. The coding corresponds to table 2. Second is a binary variable registering the presence of a volunteer donor group in a country. Individuals from countries with such groups (Denmark, France, Greece, Italy, and Spain) are coded "1," others are coded "0." Third is the presence of a commercial plasma operation (where individual suppliers are paid for plasma). Because only two countries (Germany and Spain) fall into this category, it is not included in the general model. Instead I examine how pooled results for these two countries differ from the average.

The individual-level variables are described in table 3. They have been coded with reference to the "modal donor" reported in the research studies described above. The network variable is constructed from a series of questions where respondents were asked to say whether they themselves had ever received a blood transfusion or whether they knew of a family member or a friend who had ever gotten one. The variables have been centered in order to give a substantive interpretation to the intercept term in the models. Age has been centered on 35 year olds and education on those with 16 years of full-time education. Income has been centered on the top quartile threshold. The intercept therefore represents a nonattending 35-year-old male in the top income quartile, with (roughly) a college education and no network ties to transfusion recipients.

I first specify a model with individual-level variables only, which I

⁷ The dependent variable for the mixed-effects model is "ever given" rather than "given in the last year" for two reasons. First, very few respondents had given blood in the previous year, which made fitting the model difficult. (It became overparameterized easily, with the addition of interaction terms and country-level random effects.) The "ever given" variable has a larger number of respondents. This is not ideal, but across countries, the two variables are highly correlated ($r = 0.91$). Second, from a substantive point of view, the two variables offer a more differentiated picture of the donor pool when taken together. It enables us to see that some systems recruit a large number of one-time donors, whereas others keep a smaller pool of regular volunteers.

TABLE 3

DESCRIPTION OF INDIVIDUAL-LEVEL VARIABLES

Variable	Description
Female	Coded "1" for female, "0" for male
Age	Age in years; centered on 35 year olds
Education	Years of full-time education; centered on respondents with 16 years of full-time education
Income	Eurobarometer's comparative income variable, ranging from 1 (lowest) to 12 (highest); this variable harmonizes income scores measured in different currencies
Attend	Measures whether the respondent regularly attends a church; values range from 0 (never attends) to 4 (attends daily)
Network	A four-category code measuring contacts with transfusion recipients (including oneself); values range from 0 (no ties) to 3 (ties to self, relative, and friend)

apply to each country in turn. Under this model, the log odds of individual i in country j ever having given blood are

$$\log \left(\frac{p_{ij}}{1 - p_{ij}} \right) = \beta_{0j} + \beta_{1j}F_{ij} + \beta_{2j}A_{ij} + \beta_{3j}E_{ij} + \beta_{4j}I_{ij} + e_j, \quad (1)$$

where p is the probability of having given blood in the past year; F , A , E , and I are scores for female, age, education, and income variables; the β s are unknown regression coefficients; and e is the error term. In this equation, the subscript j on the β s indicates that the individual effects are allowed to vary by country (i.e., we specify a separate regression for each country). This model allows us to see the extent of individual-level variation across countries.

For a stronger test of the institutional effects, I estimate a generalized linear mixed-effects model for the pooled sample. The institutional variables for collection regime and donor group are included and interacted with the individual-level variables. The interaction effects can be interpreted as showing how different institutional mechanisms modify the effect of individual characteristics on donation.

A standard logistic model assumes that each observation is sampled independently. This is not appropriate for a pooled sample where responses are clustered by country. The mixed-effects model used here specifies a country-level random effect to account for the fact that observations from the same country are not properly independent of one another. Re-

sults will tend to be more conservative than in the usual regression model. The log odds of individual i in country j ever having given blood are

$$\log \left(\frac{p_{ij}}{1 - p_{ij}} \right) = X_{ij}\beta + \alpha_j, \quad (2)$$

where X is a vector of independent variables and α is a random effect that varies across countries. Substantively, the random effect accounts for unmeasured factors associated with each country, thereby providing a more conservative estimate of the institutional-level variables, which vary across countries.

RESULTS

Individual Variation by Country

Table 4 shows the results from a series of logistic regressions where five individual-level variables were regressed on the donor variable (had the respondent given blood in the last year) for each country in the sample. Countries are grouped by regime type. The variables test predictions about individual donors (hypotheses 1–4 above) and also indicate the range of cross-national variation in the donor profile.

Hypothesis 1 is broadly confirmed by the data. Men everywhere are more likely to donate than women, and the odds of donating decline with age. The uniformly negative age coefficients indicate that each additional year after the age of 35 reduces one's probability of being a donor by between 2% and 6%. This is especially the case in state-run systems. The gender gap in donation is quite variable. Women may be as little as 3% or 4%, and as much as 70%, less likely to donate, depending on the country involved. (I return to this variability when I discuss hypothesis 8, below.)

Hypotheses 2 and 3 predict that the odds of donating increase with educational attainment and income, respectively. These claims do not seem to be supported by the data. France and Norway are the only two countries where higher educational attainment makes donation more likely. In all of the Red Cross countries, the education coefficients are actually negative (though not significant). Higher income significantly raises the probability of donation in five countries in the sample (Ireland, Belgium, Denmark, Norway, and Portugal).

Does the absence of strong education and income effects imply that previous studies are in error? An explanation is that while giving blood is indeed positively associated with these variables, very few people have given blood in the past year. If we regress the same independent variables in table 4 on a measure of whether the respondent has *ever* given blood,

TABLE 4

LOGISTIC REGRESSION ON DONOR VARIABLE: INDIVIDUAL-LEVEL EFFECTS BY COUNTRY

Country	Intercept	Female	Age	Education	Income	Network	Attend	N
National systems:								
Britain	-2.506*** (-4.08)	-.123 (-.54)	-.046*** (-3.04)	.017 (.17)	.007 (.10)	-.094 (-.29)	.155 (.75)	383
France	-1.620*** (-5.39)	-.064 (-.43)	-.027** (-2.45)	.142*** (2.97)	.055 (1.07)	.172 (.85)	.034 (.23)	527
Ireland	-3.331*** (-4.13)	-.334* (-1.75)	-.048*** (-3.03)	-.173 (-1.53)	.130* (1.89)	.053 (.17)	.154 (.72)	475
Red Cross systems:								
Belgium	-4.243*** (-4.70)	-.354 (-1.13)	-.024 (-1.17)	-.018 (-.15)	.206** (1.98)	.166 (.39)	.749* (2.28)	341
Luxembourg	-4.363*** (-5.60)	-.888** (-2.70)	-.026 (-1.30)	-.088 (-1.10)	.121 (1.43)	-.106 (-.31)	.660* (2.41)	314
The Netherlands	-3.270*** (-5.93)	-.200 (-1.03)	-.027* (-1.93)	-.028 (-.52)	.093 (1.55)	.502* (1.86)	.417* (2.18)	347
Germany	-3.737*** (-9.27)	-.417** (-2.67)	-.029** (-2.79)	-.023 (-.56)	.056 (1.05)	.730*** (3.47)	.285* (1.77)	907
Banking systems:								
Denmark	-1.918*** (-5.80)	-.063 (-.43)	-.019* (-1.82)	-.022 (-.64)	.104*** (2.08)	.090 (-.44)	-.049 (-.25)	561
Norway	-1.86*** (-4.37)	-.033 (-.15)	-.023 (-1.47)	.042* (1.64)	.159** (2.21)	-.015 (-.05)	-.432 (-1.48)	469
Italy	-3.172*** (-4.90)	-.593** (-2.74)	-.042*** (-3.18)	-.027 (-.55)	-.017 (-.25)	-.409 (-1.27)	.324 (1.50)	566
Greece	-2.06*** (-4.22)	-1.022*** (-5.69)	-.047*** (-4.64)	.026 (.75)	.064 (1.16)	.108 (.45)	.042 (.21)	687
Spain	-2.499*** (-5.31)	-.106 (-.55)	-.027* (-1.91)	.040 (.83)	.022 (.34)	.226 (.74)	.0271 (.16)	477
Portugal	-4.629*** (-5.95)	-.806*** (-3.18)	-.013 (-.85)	-.011 (-.22)	.168** (1.98)	.834** (2.01)	.211 (.88)	718

NOTE.—t-values are given in parentheses below coefficients.

* $P < .1$.** $P < .05$.*** $P < .01$.

then we find both income and education have positive, significant effects in most countries (though their effects remain weaker under Red Cross systems).⁸

Hypothesis 4 predicts that knowing people who have received a transfusion has a positive effect on the odds of donating. This is true for the Netherlands, Germany, and Portugal. Respondents from these countries who knew one or more transfusion recipients were between 1.6 and 2.3 times as likely to have given blood in the previous year. As with income and education, however, this effect is greatly increased if we look at those who have ever given blood rather than those who have given recently. In this case, having a network connection to a transfusion recipient is significant in seven of the thirteen countries. While knowing a transfusion recipient may provoke you to give blood (perhaps around the time of their operation), it is not nearly as likely to turn you into a regular donor.

Blood Donation and Regular Volunteering

It takes some effort to go and give blood. As such, donation should be positively related to similar activities more than to general attributes such as educational attainment. Hypotheses 5 and 6 link blood donation to involvement in religious organizations and other kinds of voluntary activity.

Hypothesis 5 predicts that countries with high rates of volunteering will also have high rates of blood donation. To test this, I compared national volunteering rates calculated from the 1991 EVS to blood donation rates calculated from the data analyzed in this article.⁹ Across all countries, there is no correlation ($r = -0.004$) between the volunteering rate and the proportion of the population who had given blood in the previous year. Contrary to the view that sees it as the perfect example of such behavior, this result suggests that blood donation is not the same sort of activity as regular volunteering at all.

The picture is subtler than this, however. Hypothesis 6 predicts that church attendance increases the likelihood of donation, just as it increases other kinds of volunteering. Here we find a surprising pattern. Regular

⁸ Results available on request.

⁹ I recalculated the rates reported in Greeley (1997) using an EVS question that asked the respondent if they did unpaid work for one of 16 kinds of voluntary organizations, ranging from church organizations and trade unions to sports, animal rights, community action, and other groups. The volunteering rate is the proportion of the population that did unpaid work for at least one of these activities. It ranges from a low of 7.6% in Spain to a high of 35.6% in Norway. The comparable statistic for the United States (from the same survey wave) is 47%.

church attendance positively and significantly raises the likelihood of donation in every country where the Red Cross runs the blood supply, but nowhere else. Regular attenders in Red Cross countries are between 1.3 and 2.1 times more likely than average to have given blood in the past year. The effect of attendance is not significant under other systems. (Indeed, it is negative in Norway and Denmark.)

This finding indicates that the Red Cross uses a collection strategy that directly or indirectly selects for religiously active people (who are also more likely to volunteer). One could interpret these results as saying that the Red Cross organizes its blood collection through religious organizations. Indeed, it is true that blood drives often target church congregations—this is an ordinary but important way in which altruism is evoked and sustained by blood collectors.

But blood banks and state systems also avail of this strategy. The Red Cross is distinctive in other ways. It is a secular organization founded on humanitarian principles, and it does not affiliate with particular religions. Each national organization maintains an extensive network of local branches that do many different kinds of social service and relief work in addition to blood collection. The Red Cross uses its own organizational structure to independently recruit people who also tend to be involved in other voluntary activities. Structurally, the Red Cross is more like other large voluntary organizations than either a state agency or a blood bank. It integrates blood donation into a wider range of voluntary activities, recruiting and keeping volunteers as it goes. This is not true of state systems, where blood collection is part of the national health service and thus institutionally isolated from other kinds of volunteering; and it can be only partly true of banking systems, which are much more limited in scope than the Red Cross and often use incentives that go against a strictly voluntary ethos.

The data support this interpretation. In Red Cross countries, the general volunteering rate (measured by the EVS) and the blood donation rate are strongly correlated ($r = 0.81$), though we saw above that there is no association between them in general.¹⁰ If we look at the proportion of donors who say they have given blood many times, we find that countries where the Red Cross has a monopoly on collection outscore all of the state systems and all but two of the banking systems. Red Cross systems maintain a pool of regular volunteers, whereas state systems adopt a more extensive approach. France and Luxembourg provide a good contrast here. France has the most donors, with 44%; Luxembourg the fewest, with

¹⁰ The corresponding correlations for the other regimes are $r = 0.28$ (state systems) and $r = -0.13$ (blood banks).

14%. But only 33% of French donors have given many times, compared to 53% of donors in Luxembourg.

Institutional Effects

The country-level results discussed so far strongly suggest that institutional variation has important effects on the blood supply. To make a stronger test of this claim, I analyzed the pooled sample in a mixed-effects model (with a random effect for each country), keeping the individual-level variables as before and adding a number of institutional-level ones. The organizational variables are interacted with the individual ones. The dependent variable is whether the respondent has ever given blood.¹¹

The results from this model are reported in table 5. The first column of coefficients shows the main effects of each of the variables in the model (reported in logits, as before). The other four columns show the interaction effects, which model the effects of the institutional conditions on the individual characteristics. Thus, -0.333 , the first number in the second column, shows the additional effect of being female under a Red Cross system.

The main finding from this model is that, as the country-level analysis strongly suggested, donation patterns in Red Cross and state regimes differ significantly from one another. For the individual effects, the female and education coefficients show that Red Cross regimes have a significantly wider gender gap than state regimes but that they reduce the effect of education. The network variable is also (and only) significant under a Red Cross regime, suggesting that the Red Cross is better able to capitalize on people's ties to transfusion recipients than its state or banking counterparts.

Hypothesis 6 receives further support from the mixed-effects model. Regular church attendance significantly increases the chances of ever having given, but only under a Red Cross regime. This again supports the claim that Red Cross regimes use collection strategies that select for regular church attenders, thus making blood donation more like other forms of volunteering.

Hypothesis 7 predicts that state-run systems will collect the most blood, followed by Red Cross systems. From our findings about volunteering, we can now see that this is an ambiguous claim. The mixed-effects model shows that state regimes will indeed have more donors in their population than Red Cross systems. The Red Cross main effect is significantly nega-

¹¹ Note that this means the interpretation of the age coefficient is different from before. It captures the now inevitably increasing effect of age on one's chances of ever having given blood.

TABLE 5

MIXED-EFFECTS MODEL OF INDIVIDUAL- AND INSTITUTIONAL-LEVEL VARIABLES ON
DONATION RATES BY COLLECTION REGIMES

Variables	Interaction Effects			
	Main Effect	Red Cross	Blood Banks	Donor Groups
Intercept111 (.34)			
Female	-.471*** (-3.70)	-.333** (-1.96)	-.402*** (-2.64)	-.305** (-2.05)
Age005 (1.14)	-.008 (-1.44)	.004 (.76)	-.003 (-.56)
Education085*** (3.41)	-.066** (-2.26)	-.052** (-2.06)	-.007 (-.42)
Income112*** (5.20)	-.044 (-1.55)	-.020 (-.78)	-.033 (-1.32)
Network127 (1.36)	.353*** (2.90)	-.097 (-.89)	.086 (.80)
Attend102 (1.61)	.101** (2.49)	-.077 (-1.29)	.028 (.36)
Red Cross	-.707* (1.67)			
Blood bank	-.320 (-.83)			
Donor group	1.044*** (2.88)			

NOTE.—Valid $N = 6,904$; scaled deviance = 7,405; scaled Pearson $\chi^2 = 6,867$; t -statistics are given in parentheses.* $P < .1$.** $P < .05$.*** $P < .01$.

tive, indicating the lower likelihood of being a donor under that system. But the dependent variable here is whether the respondent has *ever* given blood. State regimes therefore have more donors in this “extensive” sense: they are more likely to have persuaded people to donate at some point in their lives. But Red Cross regimes have more regular donors. They represent a more “intensive” strategy: although they recruit fewer donors, they are better at keeping them on as regular volunteers.

Hypothesis 8 predicts that blood banks will show the widest range of variation in donation rates. The intercept terms in table 4 bear this out. State regimes are clustered amongst the higher values; Red Cross regimes are amongst the lower. (The significantly negative main effect of the Red Cross variable in the mixed effects model reflects this clustering.) But both the lowest (Portugal) and second-highest (Norway) values are found in

banking systems. Denmark and Norway have the highest number of donors who say they have given many times. The other banking systems are at the bottom of the scale.

This variation is most obvious for gender. Italy, Greece, and Portugal show particularly large differences between men and women. Italian women are only about a third as likely as their male counterparts to have donated within the previous year. In Greece and Portugal, women are only 40% to 60% as likely as men to have donated. But, again, Denmark and Norway have no significant gender gap. Banking regimes are split on a geographical axis, with the Scandinavian systems doing well and the Mediterranean ones doing badly. This suggests that broader (perhaps cultural) features of these societies are responsible for the wide gender gap.

Hypothesis 9 predicts that where a commercial plasma supply exists the poor will be more likely to sell their plasma than donate their blood. This claim is difficult to test given the data, but there is some support for it. We have seen that income has a broadly positive effect on donation. If we separate out respondents in the bottom income quartile, we find that they are everywhere less likely to donate their blood than those with higher incomes. In Spain and Germany, the two countries where it is possible to sell one's plasma, this effect is larger than average. The opportunity to sell plasma does reduce one's likelihood of giving blood. However, the data do not allow us to say conclusively whether the poor sell their plasma more than the rich: too few respondents have ever given plasma, and we cannot distinguish between those who sold it and those who donated it.

Finally, hypothesis 10 predicts that donor organizations increase the likelihood of donation, but only for those who are already likely to give. We find that respondents were more than 2.8 times as likely to have ever donated if a donor organization existed in their country. We also find that where donor groups exist, women are 25% less likely to have given, over and above the other individual and institutional variables controlled for in the model. But the presence of a donor group did not significantly change the effect of the other individual variables. Thus, donor groups do significantly increase the number of donors, but they also make it more likely that the new recruits will be men rather than women.

DISCUSSION

It is easy to see why blood donation is thought of as an exemplary act of altruism. What could be more selfless than giving away one's own blood to a stranger in need? Yet it is an odd kind of gift, for it cannot be given away as we might hand money to a needy stranger. It must instead be

collected by a specialized organization and distributed on our behalf. In a way, this only increases the altruistic significance of donation, as it means donors cannot know or be thanked by those who receive their blood. But it also makes blood collection an organizational problem rather than an individual action. With the exception of Titmuss's pioneering effort of 30 years ago, the role of institutions in producing volunteer donors has not been studied comparatively.

This article outlined the range of organizational variation in European blood collection and showed its effects on donation. We find that state systems have a larger than average donor base, concentrated in a male population of relatively high socioeconomic status. Countries with Red Cross systems, by contrast, tend to have fewer donors, and educational attainment and income have less of an effect. Countries with a blood-banking system show the widest range of variation in donation rates. Volunteer donor groups increase the size of the donor pool but recruit more men than women as they do so.

The analysis also revealed a surprising pattern with broader implications for our understanding of volunteering and altruism in general. Collection regimes do not simply increase or decrease the donation rate along a sliding scale. They shape the kind of activity that blood donation is. The analysis suggests that under a state regime, blood donation is something that people are likely to do once or twice, probably when they are students or during a large collection effort. Under a Red Cross regime, blood donation is coordinated by a large voluntary organization that recruits a smaller group of regular donors. This more committed donor pool is likely to be made up of church-attending people who are involved in other voluntary activities. State systems pick up donors more extensively but do not retain them as successfully. Nor do they tap into the potential relationship between blood donation and other forms of volunteering or the ties of donors to those in need of blood. Under Red Cross systems, then, blood donation really is the exemplar of altruism it has always been thought to be—but this is because the collection regime has made it so, not because of the act itself or the strictly individual qualities of donors.

Future research should attend to how collection regimes influence the amount of blood collected and to how their collection strategies affect the meaning of donation. It is surprising that so little systematic data is available on the amount of blood collected each year in Europe or the United States. There is very little monitoring of the procurement or traffic in blood and blood products, despite its medical and social value (General Accounting Office 1999, p. 3). Note also that this lack of data makes it difficult to address the relationship between blood collection and blood usage. This article examined the "supply side" of the blood market. Blood is difficult to collect, so supply bottlenecks might determine usage levels.

But it might also be that medical conventions for blood use vary by country and that demand for blood varies along with them. Further research is needed here.

As for blood donation and volunteering, more qualitative data is required from specific countries. Rich information on particular regimes has yet to be collected. This line of research would carry the valid kernel of Titmuss's project into the largely noncommercial contemporary environment. Such data would allow us to further investigate the link between blood donation and volunteering identified in this article, as well as to examine the place of similar products such as human organs and genetic material. The organizational approach presented here should also be complemented by an analysis of the cultural work involved in exchanging goods of this sort, both societally and in specific interactions (Zelizer 1985, in press; Espeland 1984).

Comparative social scientists have long been aware that markets of all kinds are deeply embedded in different national contexts and that these differences can drive variation in key economic variables. This article shows that we can apply this insight to altruistic as well as self-interested exchange. How you organize a blood supply system not only affects how much you collect and who you get it from, it shapes the character of donation. Collection regimes can make blood donation work like other forms of giving, or they can make it quite unlikely that a donor will give more than once. Individuals may be moved to give their blood for any number of reasons, but it is the collection regimes that give individuals the chance to donate in the first place. Collection regimes embed altruism by creating opportunities to give. In the process, they produce differing donor populations and show us that there is more than one way for a society to rely on the kindness of strangers.

REFERENCES

- Chaves, Mark. 1998. "The Religious Ethic and the Spirit of Nonprofit Entrepreneurship." Pp. 47–65 in *Private Action and the Public Good*, edited by Walter W. Powell and Elisabeth S. Clemens. New Haven, Conn.: Yale University Press.
- Clotfelter, C. 1993. "On Trends in Private Sources of Support for the US Non-Profit Sector." *Voluntas* 4 (2): 190–95.
- DiMaggio, Paul, and Helmut Anheier. 1990. "The Sociology of Nonprofit Organizations and Sectors." *Annual Review of Sociology* 16:137–59.
- Drake, A. W., S. N. Finkelstein, and H. M. Sapolsky. 1982. *The American Blood Supply*. Cambridge, Mass.: MIT Press.
- Elster, Jon. 1990. "Selfishness and Altruism." Pp. 44–53 in *Beyond Self-Interest*, edited by Jane J. Mansbridge. Chicago: University of Chicago Press.
- Espeland, Wendy. 1984. "Blood and Money: Exploiting the Embodied Self." Pp. 131–55 in *The Existential Self in Society*, edited by Joseph A. Kotarba and Andrea Fontana. Chicago: University of Chicago Press.

- Etzioni, Amitai. 1988. *The Moral Dimension: Toward a New Economics*. New York: Free Press.
- General Accounting Office. 1999. *Blood Supply: Availability of Blood to Meet the Nation's Requirements* (GAO/HEHS-99-187R). Washington, D.C.: General Accounting Office Press.
- Genetet, Bernard. 1998. *Blood Transfusion: Half a Century of Contribution by the Council of Europe*. Strasbourg: Council of Europe.
- Granovetter, Mark. 1985. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91:481-510.
- Greeley, A. M. 1997. "The Other Civic America: Religion and Social Capital." *The American Prospect* 32 (May-June): 68-73.
- Hagen, Piet J. 1993. *Blood Transfusion in Europe: A "White Paper"*. Strasbourg: Council of Europe.
- Healy, Kieran. 1999. "The Emergence of HIV in the U.S. Blood Supply: Organizations, Obligations and the Management of Uncertainty." *Theory and Society* 28: 529-58.
- Hodgkinson, V., and M. Weitzman. 1992. *Giving and Volunteering in the United States: Findings from a National Survey*. Washington D.C.: Independent Sector.
- Jencks, Christopher. 1987. "Who Gives to What?" Pp. 321-39 in *The Nonprofit Sector: A Research Handbook*, edited by Walter W. Powell. New Haven, Conn.: Yale University Press.
- Lee, Lichang, Jane Allyn Piliavin, and Vaughn R. A. Call. 1999. "Giving Time, Money and Blood: Similarities and Differences." *Social Psychology Quarterly* 62: 276-90.
- LeGrand, Julian. 1997. "Afterword." Pp. 333-39 in Richard Titmuss, *The Gift Relationship: From Human Blood to Social Policy*, expanded and updated ed. Edited by Ann Oakley and John Ashton. New York: New Press.
- London, P., and B. M. Hemphill. 1965. "The Motivations of Blood Donors." *Transfusion* 14:559-68.
- Oswalt, R. M. 1977. "A Review of Blood Donor Motivation and Recruitment." *Transfusion* 17:123-35.
- Oswalt, R. M., and T. E. Hoff. 1975. "The Motivations of Blood Donors and Non-Donors: A Community Survey." *Transfusion* 15:68-72.
- Perlmutter, Ted. 1991. "Italy: Why No Voluntary Sector?" Pp. 157-88 in *Between States and Markets: The Voluntary Sector in Comparative Perspective*, edited by Robert Wuthnow. Princeton, N.J.: Princeton University Press.
- Piliavin, J. A. 1990. "Why Do They Give the Gift of Life? A Review of Research on Blood Donors since 1977." *Transfusion* 30:444-59.
- Piliavin, Jane Allyn, and Peter L. Callero. 1991. *Giving Blood: The Development of an Altruistic Identity*. Baltimore: Johns Hopkins University Press.
- Piliavin, Jane Allyn, and Hong-Wen Charng. 1990. "Altruism: A Review of Recent Theory and Research." *Annual Review of Sociology* 16:27-65.
- Radin, Margaret. 1996. *Contested Commodities*. Cambridge, Mass.: Harvard University Press.
- Reif, Karlheinz, and Eric Marlier. 1996. *Eurobarometer 41.0 (MRDF)*. Ann Arbor, Mich.: Inter-University Consortium for Political Research.
- Roberts, R. D., and M. J. Wolkoff. 1988. "Improving the Quality of Whole-Blood Supply: Limits to Voluntary Arrangements." *Journal of Health Politics, Policy and Law* 13:167-78.
- Salamon, Lester, and Helmut Anheier. 1998. "Towards a Comparative Theory of the Nonprofit Sector." Paper presented at the Annual Meetings of the American Sociological Association, San Francisco, August.
- Titmuss, Richard. 1971. *The Gift Relationship: From Human Blood to Social Policy*. New York: Vintage.

- . 1997. *The Gift Relationship: From Human Blood to Social Policy*, expanded and updated ed. Edited by Ann Oakley and John Ashton. New York: New Press.
- van Aken, W. G. 1993. *The Collection and Use of Human Blood and Plasma in Europe*. Strasbourg: Council of Europe.
- Walzer, Michael. 1983. *Spheres of Justice*. New York: Basic Books.
- Wuthnow, Robert. 1991a. *Acts of Compassion: Caring for Others and Helping Ourselves*. Princeton, N.J.: Princeton University Press.
- Wuthnow, Robert, ed. 1991b. *Between States and Markets: the Voluntary Sector in Comparative Perspective*. Princeton, N.J.: Princeton University Press.
- Zelizer, Viviana. 1985. *Pricing the Priceless Child: The Changing Social Value of Children*. New York: Basic Books.
- . Forthcoming. "The Purchase of Intimacy." *Law and Social Inquiry*.

The Evolution of Sex Segregation Regimes¹

Mariko Lin Chang
Harvard University

This article addresses issues of cross-national convergence in patterns of occupational sex segregation in the context of a new typology that distinguishes between substantive-egalitarian, formal-egalitarian, traditional family-centered, and economy-centered systems. Each of these systems can be characterized by distinct underlying gender "logics" and by the context of state response to issues of gender equality in the labor market. Using census and labor force survey data from 1960 to 1990 for 14 industrialized countries, log-linear models are employed to evaluate how levels and patterns of occupational sex segregation have evolved over this time period. Analyses reveal that cross-national variation in both the levels and patterns of segregation is declining over time; but at the same time, the remaining diversity among countries is increasingly patterned according to one of four segregation regimes. It appears that wholly idiosyncratic cross-national differences in the contours of occupational sex segregation are withering away as countries come to settle on, with ever-fewer exceptions, one of four possible segregation regimes.

The long-standing presumption has been that occupations are the backbone of the *class* stratification system, but as women enter into the formal economy in ever-increasing numbers, the occupational structure becomes the main locus of *gender* stratification as well. That is, occupational segregation affects the gender gap in earnings, the likelihood of career mobility, and the possibility of having some degree of autonomy over one's work or of exercising authority over the work of others (see Reskin [1993] for a review).

¹ An earlier version of this article was presented at the 1996 Annual Meetings of the American Sociological Association in New York. This research was supported by National Science Foundation (Grant SBR-9501687). My greatest debt and deepest gratitude is offered to David Grusky. I also wish to thank Maria Charles, Joon Han, Lowell Hargens, Judith Koons, John Meyer, Francisco Ramirez, Barbara Reskin, Szonja Szelenyi, Nancy Tuma, Kim Weeden, Harold Wilensky, and the *AJS* reviewers. Each contributed in distinct and significant ways to the improvement of the manuscript. Direct correspondence to Mariko Chang, Department of Sociology, Harvard University, Cambridge, Massachusetts 02138. E-mail: mchang@wjh.harvard.edu

© 2000 by The University of Chicago. All rights reserved.
0002-9602/2000/10506-0007\$02.50

Even when segregation is less consequential for occupation-related rewards, such as earnings or prestige, occupational segregation is in and of itself consequential for women's socioeconomic status. First of all, jobs filled primarily by women have a tendency to become ghettoized (Reskin and Roos 1990). Second, if women choose occupations based on family considerations *or* if a large presence of women in an occupation renders that occupation more "friendly" to workers with family responsibilities, then the concentration of women into these types of occupations may reinforce a gendered division of labor in which women's (but not men's) labor force decisions are disproportionately influenced by family and household responsibilities. Occupational sex segregation "may be a symptom and a cause of continuing gender inequality" (Rubery, Fagan, and Maier 1996, p. 431).

This article provides new evidence on the cross-national structure of occupational sex segregation. It departs from prior research both with respect to the methods deployed and the conceptual framework that is adopted. On matters of method, it should be stressed that the existing literature on cross-national sex segregation tends to rely on a scalar index, such as the index of dissimilarity, to represent the level of occupational sex segregation for each country. While such measures provide convenient snapshots of the distribution of men and women across occupations, at best they leave many interesting issues unanswered, and at worst they provide misleading portrayals of women's economic status across societies.

For example, while the United States and Japan have similar levels of occupational sex segregation, the distribution of men and women across occupations varies in ways that have important implications for the economic position of women in both countries (Brinton 1993; Brinton and Ngo 1993; Chang, Charles, and Han, in press). In Japan, men and women are more evenly distributed across the lower-status, lower-paying production occupations, yet very high levels of segregation exist in the higher status occupational groups. By contrast, the opposite pattern holds in the United States, where managerial occupations are more integrated than production occupations. In order to understand how segregation is related to gender equality more broadly, it is important to know *where* women are concentrated in the occupational structure and *how* occupational segregation fits within the larger context of gender stratification. The models presented below will make it possible to characterize cross-national and over-time variability in the patterning of sex segregation and relate such variability to macrolevel institutional forms.

The main conceptual contribution of this article is that it introduces a typology that links state gender policy and institutions to patterns of segregation. The typology will be used as a framework for examining

cross-national differences in the context of occupational sex segregation and the evolution of these cross-national differences over time. After introducing this typology, I shall examine whether patterns of occupational sex segregation across these 14 countries are (a) converging toward a common regime, (b) converging toward one of the four types represented in the typology (while wholly idiosyncratic arrangements gradually wither away), or (c) maintaining either unchanging or increasing diversity. Such a holistic and dynamic approach is critical for examining the ways in which the state, the family, and the labor market are interconnected and generate distinctive segregation profiles.

ARGUMENTS FOR THE CONVERGENCE OF STRATIFICATION OUTCOMES

With few exceptions (e.g., Anker 1998; Hakim 1993; Jacobs and Lim 1992; Rubery and Fagan 1995), existing cross-nationally comparative research on occupational sex segregation has been largely static and has accordingly failed to examine how levels of segregation may be changing over time. The few studies that have addressed cross-national trends in occupational sex segregation do not consider whether such trends imply "convergence."²

There is much speculation and theorizing about convergence in modern stratification systems (e.g., Erikson and Goldthorpe 1992; Grusky and Hauser 1984). These theories posit two explanations for why convergence in modern stratification systems is expected. First, the theory of industrialism and other structural-functional theories assert that requirements of industrial society necessitate the rise of universalistic mechanisms for allocating people to jobs (Kerr, Dunlop, Harbison, and Myers [1960] 1994). As educational and technical credentials become increasingly important in certifying people for jobs, and as bureaucratic personnel systems emerge, jobs should increasingly become allocated on the basis of merit, achievement, and credentials rather than on ascriptive criteria, such as gender. Similarly, as countries industrialize and the demand for skilled labor overshadows the demand for physical labor, job distinctions based on gender should continue to break down.

The second theoretical position that advocates convergence can be found in various cultural models that point to the autonomous rise of

² It is important to distinguish between "trends" and "convergence." While research on "trends" provide information on whether segregation is increasing or decreasing across countries, in this article the term "convergence" is used to denote an actual decline over time in the amount of cross-national variability in occupational sex segregation. This distinction is elaborated below.

egalitarian values as a mechanism that encourages convergence of stratification outcomes independent of economic change (Meyer 1994; Parsons [1970] 1994). Worldwide, there is no doubt that pervasive and growing attention is focused on issues of women's equality. In addition to various symbolic gestures, such as the recent Beijing Conference on Women, many countries have enacted antidiscrimination and equal opportunity legislation to open up opportunities for women within the paid labor market (Steinberg and Cook 1988). Women's access to higher education (Bradley 1994; Ramirez and Cha 1990) and into upper-echelon occupations (Ramirez and Weiss 1979) has also increased cross-nationally. Insofar as such changes are worldwide in nature and diffuse across nations of varied levels of development, the case can surely be made that this cultural egalitarianism is not always structural-functional in origin.

These theories leave open the possibility of two distinct forms of change. The first is a "parallel trend" in which mean levels of segregation decline (or increase) in all countries, but the extent of cross-national variability in levels of gender segregation remains fairly constant. Second, one might find "convergence" involving an actual decline over time in the amount of cross-national variability in sex segregation.³

ARGUMENTS FOR THE CONTINUED DIVERSITY OF STRATIFICATION OUTCOMES

Although theories of convergence have long been popular, there is a growing body of research suggesting a substantial degree of cross-national diversity with respect to stratification outcomes. For example, Esping-Andersen (1990, 1999) put forth a typology of three welfare state regimes, each with a qualitatively different arrangement between the state, the market, and the family. He argues that these differences are critical in shaping stratification outcomes generally, thereby precluding convergence of the sort conventionally expected by modernization theorists. For example, the "corporatist-statist" regime is active in shaping the economic and cultural subsystems, with state intervention typically maintaining class and status differences. This type of regime is often shaped by the church and is strongly committed to preserving the traditional family unit. The second type, the "social-democratic" regime, is dedicated to promoting equality, particularly along class lines. This regime is often corporatist, with interest groups occupying powerful social positions in the state decision-making process. The third type is the "liberal" regime in which the state takes a

³ It is also possible for countries to "converge" toward similar patterns of segregation without experiencing any declines in levels of segregation.

more laissez-faire approach to regulating the marketplace, intervening only in cases of market failure.

While this typology has been quite successful in explaining class-based stratification (Esping-Andersen 1990, 1993; Kolberg 1991), it is not readily applicable to systems of gender stratification and offers little explanation for cross-national differences in the nature of occupational sex segregation (Lewis 1992; O'Connor 1993; Orloff 1993; Quadagno 1988; Sainsbury 1996). For example, the United States (a "liberal" country) has taken a minimalist role in mitigating class-based inequalities, but it has been radically interventionist with respect to matters of gender and race. From the point of such class-based typologies, it may well be surprising to find affirmative action policies and quotas in the very same state that has failed to intervene in the marketplace to address class inequalities.

A TYPOLOGY OF SEGREGATION REGIMES

Although some scholars have recently sought to create typologies pertaining to gender stratification more specifically (Gardiner 1997; Lewis 1992; O'Connor 1993; Sainsbury 1995, 1996), such frameworks have not addressed differences in patterns of occupational sex segregation in detail, nor have they addressed the issue of cross-national convergence outside of the European Union. This article complements existing research by presenting a framework for investigating the connection between state policies, the family, and patterns of occupational sex segregation.

I argue that to understand cross-national variation in patterns of occupational sex segregation one must recognize the institutional context within which these sex segregation regimes are embedded and the potential role of the state in *mediating* the effects of market and family relations on women's economic status. The institutional environment clearly influences the nature of state intervention and the state's legal commitment to gender equality within the home and the workplace.

At the most abstract level, states have adopted two different approaches to issues of gender stratification: an interventionist approach in which they take an active role in shaping women's roles within the family and the formal economy, and a noninterventionist (or laissez-faire) approach in which they do not actively regulate women's role within the family or the labor force, thereby allowing other institutions (e.g., the family, the economy) to play a more dominant structuring role.

There are primarily two domains where state intervention into the nature of women's labor force participation is most likely to take place: (1) states can intervene in the public sphere by passing legislation that either promotes or inhibits women's access to participate in all occupations, and (2) states can intervene in the private sphere by taking over some of the

responsibilities of the family, such as child care, or by supplementing the income of families with children. This first form of intervention concerns "equality of access" to all occupations, and the second concerns the availability of "substantive benefits" that facilitate the employment of women with family responsibilities.

Equality of access.—In the first form of interventionism, the state may attempt to enhance women's economic status by passing various laws that promote women's equality within the labor force. Equal pay and antidiscrimination laws help to affirm women's rights to participate in the labor force on the same terms as their male counterparts.⁴ By helping to eliminate discrimination based on gender, these laws should prevent employers from restricting entrance into occupations on the basis of sex and serve as the legal backbone for workers "demands" for equality. The strongest forms of such interventionism are proactive affirmative action policies that mandate a more equal representation of men and women across occupations.

While many equal opportunity and affirmative action laws are designed to open up opportunities for women in the labor market, state intervention also comes in the form of "protective" legislation, which inhibits women from doing certain types of work because it is deemed unsafe or otherwise unsuitable for women. Laws prohibiting women from "night work" and from working in underground mines are common examples of protective legislation (ILO 1987).

Substantive benefits.—A second form of state intervention is the provision of services for working mothers that facilitate the combination of work and motherhood. There is a wide variety of benefits of this type, with the most common being guaranteed maternity leave.⁵ Less common, but perhaps even more important, is state-subsidized child care. These policies should facilitate women's employment generally and help maintain the continuous and often full-time labor force participation deemed necessary for employees in many (often male-dominated) occupations.

I propose that cross-national variation in the nature of equal access and substantive benefits result in four distinct regimes of occupational sex segregation: formal-egalitarian, substantive-egalitarian, traditional family-centered, and economy-centered systems (see fig. 1). I shall briefly review each in turn.

⁴ Research by Whitehouse (1992) suggests that the effectiveness of such legislative measures depends on the larger context of social institutions and policies.

⁵ See Gornick, Meyer, and Ross (1997) and Fagan and Rubery (1996a) for detailed discussions of how the length of maternity leave, the existence of unpaid leave, and the infrequency of fathers taking parental leave influence women's labor force attachment and opportunities.

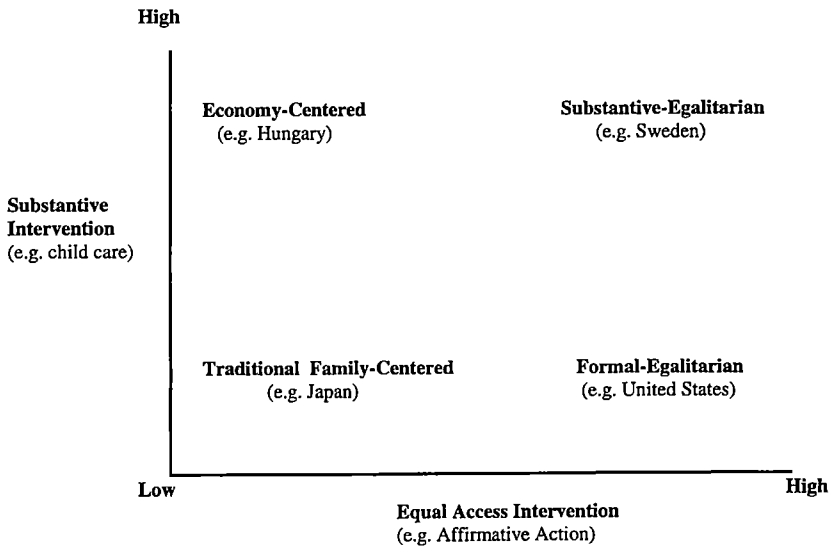


FIG. 1.—Varieties of state intervention targeted at reducing levels of occupational sex segregation.

Formal egalitarian.—Consider, first, a gender stratification regime with a formal commitment to gender equality in the labor market. As noted above, this commitment often takes the form of equal pay and antidiscrimination laws, but is most strongly evidenced by laws governing market outcomes, such as affirmative action or government quotas. Despite the formal commitment to gender equality in this type of regime, state-sponsored services that help provide women with access to paid work (such as child care) are either nonexistent or only available to small subgroups of women, usually on a means-tested basis. In this type of regime, women workers are treated as individuals rather than as mothers or potential mothers and are therefore expected to manage the complexities involved in combining work and child rearing without the help of the state.

The United States is the best example of this type. Like many countries, the United States has passed equal-pay and antidiscrimination laws. But unlike most “liberal” countries, the United States is decidedly interventionist with respect to gender inequality in the marketplace. The Equal Pay Act of 1963 and Title VII of the Civil Rights Act of 1964, as amended by the Equal Employment Opportunities Act of 1972, give the federal courts the authority to order employers to develop affirmative action programs in order to eliminate the continuing effects of past discrimination. Also, Executive Order 11246 (1971) gives the federal government the au-

thority to force employers with government contracts to adopt affirmative action programs (Christensen 1988; Reskin 1993).

Substantive egalitarian.—Alternatively, the state may accommodate demands from working women for services that support the integration of both work and family, such as state-financed child care, guaranteed parental leaves, and benefits for part-time workers. In this type of regime, a formal commitment to gender equality is coupled with a strong commitment to providing substantive support services for working women, particularly working mothers. Sweden, for instance, has state-subsidized child care, guaranteed parental leaves, and working mothers have the option of reducing their working hours while remaining eligible for work-related benefits (Kamerman and Kahn 1991; Ruggie 1984).

Traditional family centered.—In the traditional family-centered regime, issues of gender equality have received little (if any) attention by the state, neither in the form of formal-legal commitments to gender equality in the labor market, nor in the form of substantive services for working women. Typically, the cultural system promotes women's roles as belonging first and foremost within the family, as wives and mothers. Insofar as there is explicit state intervention regarding gender relations, the system of entitlements (such as the structure of pension eligibility requirements) encourages women to either remain at home or to structure their employment around family obligations.

In Japan, for example, there is a mass exodus of women from the labor force at the point of marriage or childbirth. Although many women return to the formal economy when their children leave home, by that time women have no hope of participating in the seniority-based promotion systems that provide the only route to managerial and other high-level jobs. The Japanese state has been relatively nonresponsive to issues of women's equality, so much so that it did not pass legislation discouraging sex discrimination in employment until 1985.

Economy centered.—The economy, rather than the family, plays a determinative role in shaping sex segregation in economy-centered systems. In many socialist countries, perennial labor shortages coupled with Marxian antifamily ideology led to state policy that supported the entry of women into the labor force and, in particular, into occupations that were facing the most severe labor shortages. Although socialist systems were strongly egalitarian, such sentiment was directed largely toward reducing class-based inequality, whereas the commitment to gender-based inequality was relatively superficial. In these types of systems, employment is seen primarily as a civil responsibility and hence economic requirements take precedence in the allocation of people to jobs. Economy-centered countries are similar to substantive-egalitarian countries in that they have many services available to working women, but they lack the commitment

to gender equality that is found in substantive-egalitarian regimes. While advanced socialist countries, such as Hungary and the Soviet Union, are best represented under this type, some of the less developed (third world) countries show similar economy-centered tendencies.

For example, the well-known tendency for women in Hungary and other formerly socialist states to be concentrated in medical occupations (e.g., doctors, dentists, pharmacists) arose because (a) the state invested heavily in medical services and hence the demand for medical personnel was strong in the 1950s and 1960s when young women were entering the labor force (in response to an overall labor shortage), (b) medical occupations were deskilled and underrewarded (relative to Western standards), thereby rendering them less attractive to men, and (c) professions in general were underrewarded relative to second-economy occupations, which were "closed" to women because they operated after hours when women had family responsibilities (Corrin 1994; Rueschemeyer and Szelenyi 1989).

The foregoing classification is inconsistent with simple convergence theorizing. That is, rather than expecting segregation regimes to become increasingly similar across nations, we instead expect continuing diversity generated by such regime-specific logics of sex segregation. At the same time, while it is conceptually useful to distinguish these types from one another, it is most accurate to conceive of these dimensions as existing on a continuum. Few countries fit discretely into one type, and over time countries may shift along one or more dimension on the continuum. This typology nevertheless serves to summarize underlying gender "logics" and ways in which the state responds to issues of gender equality in the labor market.

TYPOLOGICAL IMPLICATIONS FOR SEX SEGREGATION

Given that the character of women's labor force participation is embedded within a larger framework, it is unlikely that significant changes in the contours of segregation would occur independently from this larger institutional environment. In their study of sex segregation within corporations, Bielby and Baron (1984, p. 52) conclude that "large and systematic reductions in gender segregation seem unlikely to occur in the absence of fundamental shocks to the social system." I agree that meaningful changes in the patterns of occupational sex segregation are extremely unlikely without intervention of some kind. While Bielby and Baron (1984) focus on organizational change, I intend to examine the effects of intervention at the level of the state. As stated previously, I believe that the most likely sources of meaningful changes in patterns of segregation arise from either economic or state interventionist sources.

Large-scale economic changes, such as changes in unemployment rates and industrialization, alter the character of employment and the gender dynamics within the labor force. While certain changes may be advantageous to women in the paid labor force, others may hamper equality. Also, without corresponding social, political, and legal support, the economically driven gains of women in the labor force may be lost as quickly as they were won. For instance, the economically driven changes under socialism in Hungary that helped integrate women into the labor force and into some previously male occupations are currently vulnerable to new economic (and political) realities that are not as "women friendly" (Funk and Mueller 1993).

The most common type of state intervention occurs in the form of legislation adopted to address gender inequality within the labor market. When opportunities for women are expanded via equal opportunity and affirmative action legislation, women are most likely to press for entrance into the more prestigious (and traditionally male-dominated) managerial and professional occupations because of the income, prestige, and autonomy associated with these occupations (Charles 1998) and because they are allocated on the basis of educational credentials rather than ascriptive criteria. As a result, I expect that countries with strong equal opportunity or affirmative action policies would experience desegregation within the professional and managerial occupations in particular.

Whereas labor market interventions address economic inequality directly, substantive interventions take a more indirect approach to gendered labor market inequalities by providing services to working women that facilitate women's dual roles as workers and mothers. Despite efforts to minimize the tension between "home" and "work," segregation may actually increase with such substantive intervention because the expansion of services for working women can create female-typed jobs (Charles 1992; Esping-Andersen 1993; Hagen and Jenson 1988; Kolberg 1991). In this regard, one ironic by-product of the strong commitment to gender equality in substantive-egalitarian countries is that segregation increases as women flow into female-typed service jobs created by the differentiation of functions, such as child care, out of the family and into the formal economy.⁶ Despite the feminization of the service sector, gender-egalitarian ideology coupled with the availability of substantive services for working mothers should enable women to enter some male-dominated occupa-

⁶ Research by Charles (1998) suggests that the greater concentration of women in service jobs is in part a by-product of a "negative" selection effect common in countries with high levels of female labor force participation: "Where more women have paid jobs, a large share of the female labor force may have extensive domestic obligations, low levels of educational capital, or both" (p. 109).

tions, particularly the more prestigious administrative and managerial occupations. Therefore, substantive-egalitarian countries are the most likely of the four regimes to exhibit *both* high levels of segregation in the service occupations *and* an integration of the professional and managerial occupations. These seemingly inconsistent trends are a characteristic of the two countervailing forces of gender egalitarianism (which facilitates women's entrance into the professional and managerial occupations) and of the expansion of the service-oriented postindustrial economic structure (which draws women into the service sector).

Segregation in economy-centered countries follows a different pattern. I expect that the conjunction of supply and demand factors and the relatively high educational attainment of women typical of many socialist economies should result in the concentration of women in professional occupations. In similar fashion, one would expect a disproportionate number of women in production occupations, since the postwar expansion of socialist production coincided with the mass entry of women into the labor force. The high levels of substantive benefits and full-employment policies bring all women into the labor force to fill positions in the economy where there are labor shortages. While substantive benefits can potentially enhance gender equality in the labor force, their effectiveness is limited unless they are coupled with strong legislation aimed to promote gender equality.

In the absence of both labor market and substantive interventions, women who enter the formal economy will most likely "choose" occupations that are considered appropriate for women, typically because they are viewed as extensions of women's natural roles as caregivers (e.g., nursing and elementary school teaching). Alternatively, women in traditional family-centered countries may be channeled into part-time or temporary jobs that are less likely to interfere with their responsibilities in the home. Whereas certain occupations (such as nursing and teaching) are viewed in most industrialized countries as "female" occupations, the types of part-time or temporary jobs that women may engage in most likely differ across countries, since they are contingent on a particular country's economic and cultural particularities. Insofar as there is any shared segregation features in traditional family-centered countries, it may well be the relative lack of gender integration in the managerial occupations. The absence of legislation promoting equality within the labor market coupled with the cultural value placed upon women's roles within the home should prevent many women from participating in these occupations, both because they are traditionally male *and* because they typically require a continuous full-time labor force commitment.

What must therefore be stressed is that the structure of change may well vary across regime types in ways that are not necessarily consistent

with conventional convergence theories. I shall begin with analyses that test for convergence across 14 industrialized countries and then examine whether the preceding typology provides a useful framework for understanding cross-national variability in occupational sex segregation.

DATA

Data for the first set of analyses come from the *Year Book of Labour Statistics*, published by the International Labour Organization (ILO 1976, 1982, 1991, 1992, 1993, 1994, 1995; ILO 1990 [retrospective ed.]). These tables contain information on the distribution of men and women across seven occupational categories,⁷ based upon the 1968 International Standard Occupational Classification (ISCO) scheme (ILO 1968).⁸ The seven occupational categories are: (1) professional, technical, and related workers, (2) administrative and managerial workers, (3) clerical and related workers, (4) sales workers, (5) service workers, (6) agricultural, animal husbandry, and forestry workers, fishermen and hunters, and (7) production and related workers, and transport equipment operators and laborers. Since the classification protocol for agricultural workers varies widely across countries, I eliminated this category from the analyses.

I examine the period from 1960 to 1990 because of the unprecedented growth of female labor force participation during this period and the corresponding change in the size and goals of the welfare state. I have limited the sample to industrialized countries since they are most likely to have the fairly well-developed state institutions necessary for the implementation of these types of social policies. After further eliminating those coun-

⁷ Full-time and part-time workers are aggregated in these tables. Research suggests that part-time workers are more segregated than full-time workers and that patterns of segregation for part-time workers vary in important ways across countries (Fagan and Rubery 1996b; Hakim 1993). Indeed, I would expect that such cross-national variation in the characteristics of part-time workers is a critical feature of a regime's "segregation profile." A joint analysis of the segregation of full-time and part-time workers would enrich the analyses presented here.

⁸ The 1960 data reported in the *ILO Year Book of Labour Statistics* are based on an earlier version of ISCO, drafted in 1958. This earlier version contained an additional category for "workers in transport and communications occupations." While some of the constituent occupations in this category would fall under "clerical" or "professional" occupations in the 1968 scheme, most of the incumbents would have been classified within the category for "production and related workers, transport equipment operators and labourers." Since ILO reports only aggregated data, I combined the 1960 category for "workers in transport and communications occupations" with that for "production and related workers, transport equipment operators, and labourers" in order to make the categories as consistent as possible from 1960 to 1990. In doing so, I have followed the lead of Jacobs and Lim (1992).

tries for which data were not available for all four time periods, the sample consisted of 14 countries.

The second set of analyses examines convergence across four countries in much greater depth, each representing one of the four regimes in the preceding typology. Data on the distribution of men and women within detailed occupational categories comes directly from national censuses for the years 1960, 1970, 1980, and 1990 for the United States (formal egalitarian), Sweden (substantive egalitarian), Japan (traditional family-centered), and Hungary (economy centered). The occupational classification schemes were made comparable by recoding them into the 3-digit 1988 International Standard Classification of Occupations.⁹ After rendering these schemes comparable both across countries and across time, 54 occupational categories remained (see app. A).¹⁰

This reclassification effort was a formidable one, but ISCO-88 (ILO 1988) provides guidelines and a detailed explanation of types of occupations within each category and how occupations should best be translated into the ISCO categories. Although by no means perfect, this method of classification provides the best possible scheme of occupational classification for the analyses.

THE CLASSIFICATION SCHEME

The 14 countries in the analyses were classified into one of the four segregation regimes on the basis of two indices. The first, a family benefit index developed by Wilensky (1990, 2000),¹¹ classifies countries according to the degree of innovation and expansiveness of their family policies during the period of 1976–82.¹² Here, two components of Wilen-

⁹ The 1988 version of ISCO improves upon prior versions by elaborating existing occupational categories, reorganizing categories to better align occupations of similar job and skill levels, and promoting "detail and clarity of information on sex composition of jobs" (ILO 1988, ISCO-88:10).

¹⁰ There are 115 occupational categories at the 3-digit level in ISCO-88 (ILO 1988). However, in order to facilitate comparison among the four countries and time points, some categories were combined while others were elaborated. The resulting occupational classification scheme is equally suited for comparison across all four countries and time points.

¹¹ Used by permission of Harold L. Wilensky from chap. 5 of his "Rich Democracies: Political Economy and Performance" (2000). Wilensky's index has also been used by Siaroff (1994) and by Gornick et al. (1997).

¹² During this time period, rates of female labor force participation increased substantially and substantive benefits to working mothers were increasingly provided. Although this measure cannot distinguish countries that instituted generous maternity and child care benefits earlier than 1976 from those that adopted generous benefits as late as 1982, in practice most of the countries scoring high on the Wilensky index had a strong commitment to substantive services prior to 1976.

TABLE 1
LEVEL OF SUBSTANTIVE BENEFITS FOR WORKING PARENTS

Country	Child Care	Maternity/ Parental Leave	Average
Austria	1	3	2
Canada	0	1	.5
Denmark	3	2	2.5
Finland	2	4	3
Germany (FRG)	1	3	2
Hungary*	3	4	3.5
Ireland†	1	1	1
Japan	0	1	.5
The Netherlands	1	1	1
New Zealand	0	1	.5
Norway	3	3	3
Portugal‡	1	3	2
Sweden	3	4	3.5
United States	1	0	.5

NOTE.—Data are derived from Wilensky's (1990, in press) family benefit index.

* Scores for Hungary were calculated using the guidelines outlined by Harold L. Wilensky (2000) and the following sources: U.S. Department of Health, Education, and Welfare (1975, 1979; U.S. Department of Health and Human Services [1983]; and Kamerman and Kahn [1981]).

† The "child care" score for Ireland was derived from Siaroff (1994) and adjusted to correspond to the original range of Wilensky's (1990, 2000) five-point scale, ranging from zero to four.

‡ "Maternity/parental leave" score for Portugal was derived from Siaroff (1994) and adjusted to correspond to the original range of Wilensky's (1990, 2000) five-point scale, ranging from zero to four. Portugal's "child care" score was calculated using the guidelines outlined by Wilensky (2000) and data from Phillips and Moss (1989) and Moss (1990).

sky's index are used: (1) the existence and length of maternity and parental leave and (2) the availability and accessibility of public daycare programs and government efforts to expand daycare. Scores for each component range from a low of 0 to a high of 4. For our purposes, countries with an average score of 2.5 on these two components (see table 1) of the index were classified as having "high" levels of substantive services (e.g., Denmark, Finland, Hungary, Norway, and Sweden), while all other countries were classified as having "low" levels of substantive services.

A second index was developed to classifying countries by the extent of legally conferred "equality of occupational access." This index comprises two related but conceptually distinct dimensions: (1) laws aimed at providing women with equal access to all occupations (i.e., integrative legislation) and (2) laws prohibiting women from certain occupations (i.e., segregative legislation).

Three types of integrative legislation comprise this dimension of the

index.¹³ First, affirmative action (also called "positive action") policies explicitly promote the integration of occupations on the basis of sex. The second indicator of a state's commitment to gender equality is the ratification of ILO Convention 111, which prohibits discrimination in employment and occupation. The third indicator is the receipt of the instrument of ratification of the United Nations Convention on the Elimination of All Forms of Discrimination against Women (CEDAW).

The ILO (1987) notes that there are two categories of protective ("segregative") legislation: (1) legislation aimed at protecting women's reproductive and maternal functions (such as prohibiting pregnant women from working with toxic substances) and (2) legislation "aimed at protecting women generally because of their sex, based on attitudes toward their capabilities and appropriate role in society" (p. 9). This latter form of protective legislation is arguably particularly discriminatory toward women and most commonly includes the prohibition of night work by women, the prohibition of women from working underground, and restrictions placed upon the manual transport of loads by women. The ratification of ILO Convention 89 (which prohibits women's night work) and 45 (which prohibits women's underground work) are used to as two indicators of protective legislation. A third form of protective legislation is the restriction of the manual transport of heavy loads: "It is argued that limiting the employment of women in certain occupations and in jobs requiring the lifting and carrying of weights over a certain limit does not take into account the capacities, preferences and abilities of individual women and therefore is discriminatory. Instead, individuals should have the opportunity to display their ability to do a particular job" (ILO 1987, p. 8).

The index of "equality of occupational access" legislation is thus composed of six measures: (1) the presence of affirmative action policies, (2) the ratification of ILO Convention 111 prohibiting discrimination against women in employment and occupation, (3) the ratification of the UN CEDAW, (4) the *absence* of ratification of ILO Convention 89 prohibiting women from night work, (5) the *absence* of ratification of ILO Convention 45 prohibiting women from working underground, and (6) the *absence* of legislation restricting the manual transport of heavy loads by women.¹⁴

¹³ Equal pay legislation does not seek to integrate women across the occupational structure and therefore was not included as a component of this index.

¹⁴ For each of the measures, a country's "score" depends on the presence or absence of the legislation by 1982, rendering the time period consistent with the Wilensky index. This date allows countries one year to ratify the UN CEDAW after it went into force. It also provides a generous deadline for the ratification of the various ILO conventions. Indeed, almost all countries who have thus far ratified the preceding ILO conventions did so by 1982. When earlier years were used as the "cut-off" date, the ranking of countries along this dimension is fairly consistent.

The first three measures aim to enhance women's participation in all occupations; the next three measures aim to restrict women's participation in certain occupations. The scoring of countries according to this index is shown in table 2. Scores range from 0 to 6, with a "6" indicating that a country's legal framework enhances women's employment in all types of occupations. Countries with the highest scores on this index include Sweden, Norway, Canada, Denmark, and the United States.

When countries are plotted along these two dimensions (see fig. 2), it is possible to discern distinct "clusters" of countries,¹⁵ which represent the four segregation regimes presented in figure 1.¹⁶ The substantive-egalitarian countries include Sweden, Norway, Denmark, and Finland; the formal-egalitarian countries include Canada and the United States; the traditional family-centered countries include Germany (FRG), Austria, Japan, the Netherlands, Ireland, New Zealand, and Portugal; and the sole economy-centered country is Hungary.

METHODOLOGY

This article utilizes the log-linear index of segregation (A) proposed by Charles and Grusky (1995) because it has the benefit of allowing one to separate *trends* in occupational sex segregation from changes in the occupational structure and changes in levels of female labor force participation.¹⁷

¹⁵ Other researchers (e.g., Castles 1993; Gardiner 1997; Gornick et al. 1997; Sainsbury 1996; Schmidt 1993; Wilensky 1983) have found similarities among countries within these clusters with respect to critical political, social, religious, historical, and economic features.

¹⁶ The boundaries distinguishing different clusters are of course discretionary, but are meant to reflect the underlying gender "logics" of state response to gendered labor market inequalities (occupational sex segregation in particular). For instance, Finland appears to occupy an intermediate position between the economy-centered regime and the substantive-egalitarian regime (which is not surprising given its historical ties to the former Soviet Union). Analyses were carried out in which Finland was categorized as an economy-centered country and revealed that the substantive conclusions presented here would be the same if Finland had been classified as an economy-centered country. However, I believe that Finland's position as a leader with regard to women's position in society (e.g., Gardiner 1997; Norris 1987; United Nations 1995) places it closer to the substantive-egalitarian regime.

¹⁷ The major disadvantage of using the index of dissimilarity for the comparison of levels of segregation either over time or across countries is that it is sensitive to changes in the relative sizes of occupations. The size standardized index of dissimilarity overcomes this drawback by giving each occupation equal weight, but it is sensitive to changes in the percentage of women in the labor force. The choice among various segregation indices has become the subject of much debate (see, e.g., Charles and Grusky 1995; Grusky and Charles 1998; Hakim 1993; Jacobs 1993; Siltanen, Jarman, and Blackburn 1995; Watts 1993, 1998; Weeden 1998). While the choice between indices is conceptually and methodologically important, it bears emphasizing that no single summary measure can adequately capture the complexities of occupational sex segregation.

TABLE 2

INDEX OF LEGALLY MANDATED EQUALITY OF OCCUPATIONAL ACCESS

COUNTRY	"ENABLING" LEGISLATION				"PROTECTIVE" ("SEGREGATIVE") LEGISLATION			
	Affirmative Action	Equal Opportunity	UN CEDAW	Night Work	Underground Work	Heavy Loads	TOTAL	
Austria	no	yes (1973)	... (1982)*	yes	yes	no	2	
Canada	yes	yes (1964)	yes (1981)	no	no	no	6	
Denmark	no	yes (1960)	... (1983)	no	no	no	4	
Finland	no	yes (1970)	... (1986)	no	yes	no	3	
Germany (FRG)	no	yes (1961)	... (1985)	no	yes	yes	3	
Hungary	no	yes (1961)	... (1980)*	no	yes	yes	2	
Ireland	no	no	... (1985)	yes	yes	yes	0	
Japan	no	no	... (1985)	no	yes	yes	1	
The Netherlands	no	yes (1973)	... (1991)	no	yes	yes	2	
New Zealand	no	... (1983)	... (1985)	no	yes	no	2	
Norway	yes	yes (1959)	yes (1981)	no	no	no	6	
Portugal	no	yes (1959)	yes (1980)	yes	yes	yes	2	
Sweden	yes	yes (1962)	yes (1980)	no	no	no	6	
United States	yes	no	no	no	no	no	4	

SOURCES.—Hevener (1983); ILO (1987, 1992); Ladin (1994); Nielsen (1983); OECD (1994, 1998); United Nations (1996).

* Austria and Hungary ratified the convention with one or more reservations and therefore are not counted toward their overall score in this index. Hungary withdrew their reservation in 1989.

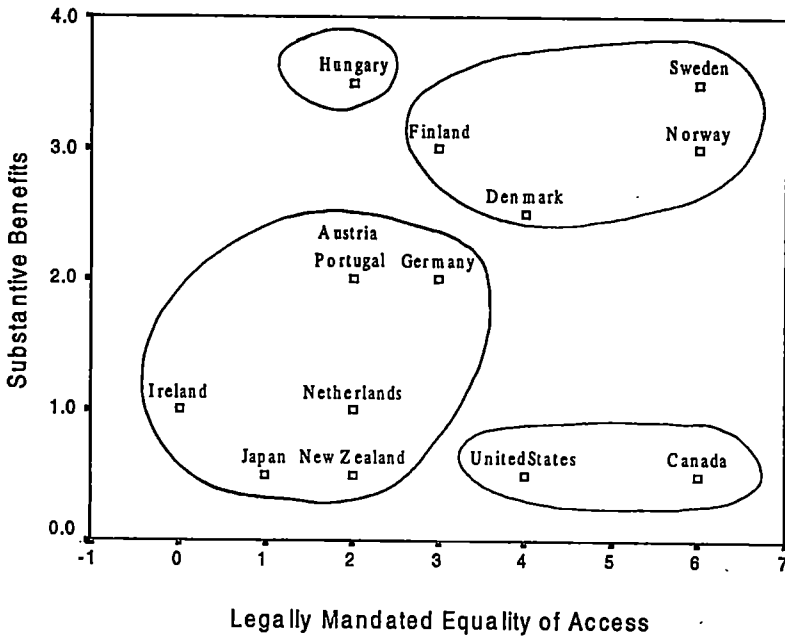


FIG. 2.—Classification of countries according to level of substantive benefits and legally mandated equality of occupational access.

This index characterizes the extent to which women are, on average, either overrepresented or underrepresented in the occupational structure net of all marginal effects. It equals the average of the squared deviations of the logged sex ratios from the mean logged sex ratio:

$$A = \exp \left\{ \frac{1}{n} \times \sum_{i=1}^n \left[\log(W_i/M_i) - \left(\frac{1}{n} \times \sum_{i=1}^n \log(W_i/M_i) \right) \right]^2 \right\}^{1/2}, \quad (1)$$

where W_i and M_i refer to the number of women and men in the i th occupation and n refers to the number of occupations. If the same sex ratio prevails in all occupations, women are equally represented throughout the occupational distribution, and A therefore equals 1. The more segregated the labor market, the higher the value of A .¹⁸

However, since any summary index may mask important differences in patterns of segregation, various log-linear models will also be used. The first log-linear model of interest addresses the question of whether the

¹⁸ In a perfectly segregated labor market, A is undefined since M_i would equal zero for all "female" occupations.

underlying structure of cross-national variability is changing over time. This model fits all three-way interactions between occupation, sex, country, and time and can be represented as $(O*S*C + O*S*T + O*C*T + S*C*T)$, where O = occupation, S = sex, C = country, and T = time. This model allows occupational sex segregation to vary across countries and over time but forces cross-national variability to assume the same pattern in all time periods. I shall refer to this specification as the model of constant cross-national variability (model 1). If this model fits the data, then the underlying structure of cross-national variability is unchanging, and the data are inconsistent with theories postulating either convergence or divergence.¹⁹

Next, I shall relax the foregoing constraints to allow the four regimes to become more or less distinctive over time. That is, whereas the prior model implies that the underlying structure of cross-national variability is unchanging, the model of constant within-regime variability (model 2) implies that cross-national variability within regimes (but not between them) is unchanging. This model can be represented as $(O*S*C + O*C*T + S*C*T + O*S*T*R)$, where R = regime type and all other symbols are defined as before. If this model can be rejected, then within-regime patterns are changing over time and a movement toward convergence or divergence is a possibility.

If the model of constant cross-national variability is rejected, then patterns of segregation are changing over time and convergence is a possibility. However, before concluding in favor of convergence theories, it must further be shown that the amount of cross-national variability declines monotonically. Therefore, I will fit a log-linear model of cross-nationally invariant segregation (model 3) for each time period that allows for an association between sex and occupation but constrains this association to be equal in all countries. This model can be represented as $(O*C + S*C + O*S)$, where all symbols are defined as before. The resulting test statistics will reveal which time periods exhibit the greatest cross-national variability.

In addition, I employ a log-linear model in each time period that allows the sex-by-occupation association to vary by regime type: $(O*C + S*C + O*S + O*S*R)$, where all symbols are defined as before. I shall refer to this specification as the model of segregation regimes (model 4).

Finally, parameter estimates for the extent of male or female overrepresentation within occupations for each country and time period will be obtained from a saturated model that fits a three-way association between occupation, sex, and country for each of the four time points. To identify

¹⁹ A lack of change in patterns of segregation over time is consistent with theories advocating persistent (and unchanging) cross-national diversity.

this model, the sex-by-occupation interaction effects have been constrained to sum to zero, with negative values indicating male overrepresentation, positive values representing female overrepresentation, and zero representing perfect integration by sex.

AGGREGATE ANALYSES

Before progressing to analyses that address convergence directly, it is informative to present an overall picture of trends in segregation from 1960 to 1990 in these 14 countries. Figure 3 displays the trends in levels of occupational sex segregation from 1960 to 1990. The results are consistent with neither a "parallel trend" nor "convergence" in the strictest sense. There is a slight decline in levels of segregation when one compares 1960 to 1990, but many countries fail to demonstrate consistent trends toward either increasing or decreasing levels of segregation across this time period.²⁰ In Austria, for example, levels of segregation decrease from 1960 to 1970, increase between 1970 and 1980, and decrease once again between 1980 and 1990. Moreover, the cross-national range and variance in values is only barely smaller in 1990 than in 1960.²¹ In 1960, the degree to which men and women were overrepresented in the average occupation ranged from a low of about 2.2 (in Hungary and Portugal) to a high of approximately 3.6 (in Sweden, Finland, Norway, and the Netherlands). By 1990, levels had declined slightly, ranging from the same low of about 2.2 to a high of approximately 3.4.

I shall next fit the model of constant cross-national variability (model 1) that allows segregation to vary across countries but constrains this three-way association to be the same across all four time periods. If model 1 fits the data, then the extent and pattern of cross-national variability in segregation is not changing over time and one can therefore conclude that there is no convergence. However, if the model fails to fit, then the saturated model is preferred and one can assume that the extent of cross-national variability in patterns of segregation is changing over time, which is consistent with convergence theories. As shown in table 3, the index of dissimilarity indicates that the model explains all but about 1% of the variation in segregation patterns. However, the likelihood and BIC statistics for model 1 reveal that this model does not fit the data, implying

²⁰ Jacobs and Lim (1992) found similar results in their analyses of 56 countries for the 1960–80 time period.

²¹ The index of dissimilarity suggests a different ranking of countries with respect to levels of segregation, but it does not reveal any change in the cross-national variance in levels of segregation nor any trends toward either increasing or decreasing levels of segregation.

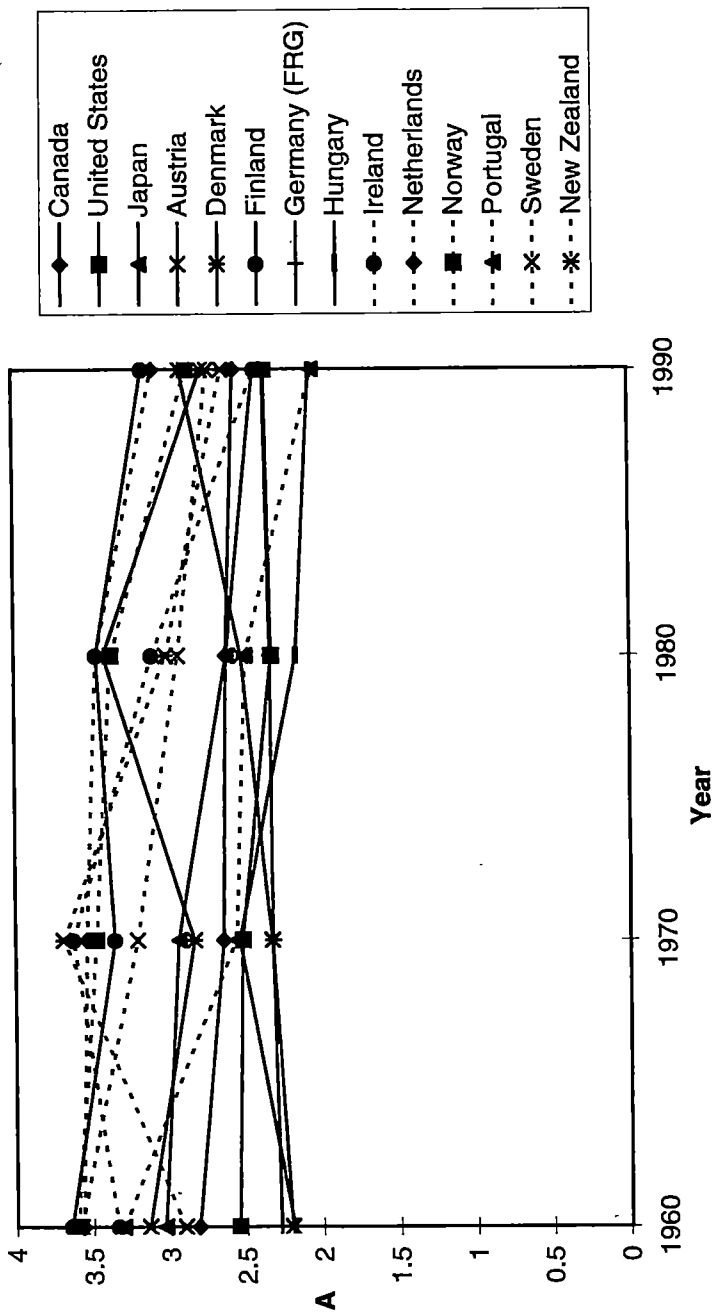


FIG. 3.—Log-linear index of segregation (A) for 14 countries, 1960–90

TABLE 3

FIT STATISTICS FROM LOG-LINEAR MODELS OF OCCUPATIONAL SEX SEGREGATION IN 14 COUNTRIES: 1960, 1970, 1980, and 1990

Model	Sample Size	L ²	df	Δ	L_i^2/L_i	BIC
1. Constant cross-national variability ($O*S*C + O*S*T + O*C*T + S*C*T$)	266,267,149	275,902	195	.95		272,119
2. Constant within-regime variability ($O*S*C + O*C*T + S*C*T + O*S*T*R$)	266,267,149	257,773	151	.87		254,803
3. Cross-national invariance ($O*C + S*C + O*S$):						
1960	91,731,520	1,873,277	65	5.1		1,872,085
1970	67,170,848	1,072,628	65	4.4		1,071,456
1980	92,698,096	1,824,947	65	5.1		1,823,754
1990	14,666,685	228,108	65	4.4		227,035
4. Segregation regimes ($O*C + S*C + O*S + O*S*R$):						
1960	91,731,520	982,425	50	3.5	52.4	981,508
1970	67,170,848	511,928	50	2.6	47.7	511,027
1980	92,698,096	370,218	50	1.5	20.3	369,301
1990	14,666,685	69,805	50	1.8	30.6	68,980

NOTE.— O = occupation; S = sex; C = country; T = time; R = regime type; Δ (the index of dissimilarity) is the percentage of cases that are misclassified under the relevant model.

that patterns of segregation are changing over time, albeit perhaps not substantially so. The total sample for the 14 countries is large enough to make it unlikely that any model save the saturated one will fit the data. Consequently, although test statistics are potentially informative, the fundamental values of interest are the parameters themselves (to which we shall turn subsequently). We shall leave it to the reader to decide whether the changes revealed by the parameters are of sociological interest.

It is also informative to examine whether the four regimes are becoming more or less distinctive over time. As argued in the introduction, one might well expect within-type idiosyncrasies to gradually disappear, whereas between-type differences might remain as strong as ever. The extent to which such within-type idiosyncrasies are indeed withering away can be assessed by fitting a model that constrains them to be of constant strength. The model of constant within-regime variability (model 2) allows for segregation to vary across countries *within each regime* but constrains this association to be the same across time periods.²² Once again, the likelihood and BIC statistics indicate that the model fails to fit the data, suggesting that the amount of within-type variability is changing over this time period.

Taken together, these results leave open the possibility that cross-national patterns of segregation may be changing over time both across all 14 countries and within regimes, but it is altogether unclear as to whether the change over time indicates *convergence*. Since it is possible to have change without convergence, I will move to a series of models that addresses the issue of convergence over time more specifically.

For each time period, I applied a log-linear model that allows for an association between sex and occupation but constrains this association to be equal in all countries (model 3). The fit statistics for this model support the conclusion that there is no convergence in patterns of segregation at the major occupational level.²³ As revealed in table 3, the percentage of misclassified individuals in model 3 (represented by the index of dissimilarity) changes over time but does not steadily decline with time. Instead, we see a decline in the index of dissimilarity from 1960 to 1970, an increase from 1970 to 1980, and a decline from 1980 to 1990.

Although the index of dissimilarity is a useful summary measure of model fit, it is affected by the sparseness or density of counts in regions of the table where the model fails to fit. It is therefore useful to examine

²² This model is equivalent to fitting the model of constant cross-national variability separately within each regime and then summing the resulting test statistics.

²³ It is important to keep in mind that the likelihood test statistic (L^2) is sensitive to sample size and therefore cannot be compared across samples insofar as the objective is to assess the extent of change in the parameter estimates.

TABLE 4

VARIANCE OF THE OCCUPATION-SPECIFIC SCALE VALUES ACROSS 14 COUNTRIES: 1960, 1970, 1980, AND 1990

Year	Professional	Managerial	Clerical	Sales	Service	Production	Average
1960146	.199	.203	.123	.105	.115	.149
1970094	.179	.155	.083	.083	.143	.123
1980052	.160	.135	.061	.140	.129	.113
1990082	.217	.113	.065	.089	.164	.122

the variance across the occupation-specific scale values from the saturated model that fits a three-way association between occupation, sex, and country for each time period.²⁴ These variances are displayed in table 4. We can see that with the exception of the clerical category, there is no consistent decline in the amount of variance in patterns of segregation from 1960 to 1990.²⁵ Once again, the results are inconsistent with a conventional convergence story.

It is natural to next ask whether the typology that I have proposed can explain this persisting diversity in segregation regimes. I estimated a model that constrains patterns of segregation to be the same for countries within each of the four segregation regimes (model of segregation regimes). By 1990, only 30.6% of the variance remains unexplained by the typology (see model 4, table 3). In other words, the model of segregation regimes accounts for approximately 70% of the variation in sex segregation across countries, with the remaining 30% occurring within the four proposed regimes.²⁶ With the exception of a slight increase in the index of dissimilarity between 1980 and 1990,²⁷ the typology accounts for an increasing percentage of the cross-national variation in patterns of segregation over

²⁴ Occupation-specific scale values (v_i) for each country and time period are derived from the following equation:

$$v_i = \ln(W_i/M_i) - \left[1/n \sum_{i=1}^n \ln(W_i/M_i) \right],$$

where all symbols are as defined in eq. (1).

²⁵ While there is some hint of convergence from 1960 to 1980, this is clearly offset by a divergent trend from 1980 to 1990. The average variance in 1990 is almost identical to the average variance in 1970.

²⁶ When Hungary was excluded from the analysis, results were consistent with those presented here.

²⁷ This slight increase in the index of dissimilarity between 1980 and 1990 is due to the increase in variance within the substantive-egalitarian regime during this period (as shall be discussed later in this section).

time, suggesting that it is residual, unexplained forms of variability that are increasingly withering away. In 1960, the typology explained roughly 50% of the variation, whereas in 1990 it explained 70%.

The underlying structure of this variability is best revealed by graphing the occupation-specific values from a saturated model that fits a three-way association between occupation, sex, and country for each of the four time points (see fig. 4). Examining trends in patterns of segregation for each regime over time, we find that the formal-egalitarian countries exhibit a dramatic decrease in segregation in the managerial category. This regime is also distinct for its high levels of segregation in the clerical and production occupations (in comparison to other regimes). Taken together, these results suggest that women as a whole are moving toward managerial and clerical occupations but away from service and production occupations. This is consistent with the claim that formal egalitarianism increases women's access to the prestigious male-dominated occupational categories.²⁸ Also, as suggested by Charles (1998), the clerical occupations in these countries become increasingly attractive to women since they are of higher prestige than blue-collar occupations.

In general, trends in the substantive-egalitarian regime parallel those found in the formal-egalitarian regime, but to a much lesser degree. While one might anticipate an increase in segregation in the service category from 1960 to 1990 (due to the postindustrial demand for service workers in the substantive-egalitarian countries, there is in fact a *decrease* in segregation. The expansion of the female-dominated service category may have been partially offset by women's movement into professional, managerial, and, to a lesser extent, clerical occupations. This is consistent with the countervailing forces outlined by Charles (1998).

Contrary to expectations, a distinct feature of the economy-centered country is the lack of women in the professional category. The occupation-specific values indicate that, whereas medical occupations within the professional category are often feminized, many of the other professional occupations are extremely male dominated, rendering the professional occupations as a whole male dominated. Hungary is also unique among the countries in the analysis in that the professional category became *increasingly* male dominated during this time period. This lack of women in the professional category is in stark contrast to the remarkable integration of the managerial category between 1980 and 1990. Although Hungary's integrated managerial category was indeed unexpected, in retrospect one has to bear in mind that the many lower-level state managers have little or

²⁸ Levels of female educational attainment are also positively associated with the presence of women in skilled white-collar occupations (Charles 1998).

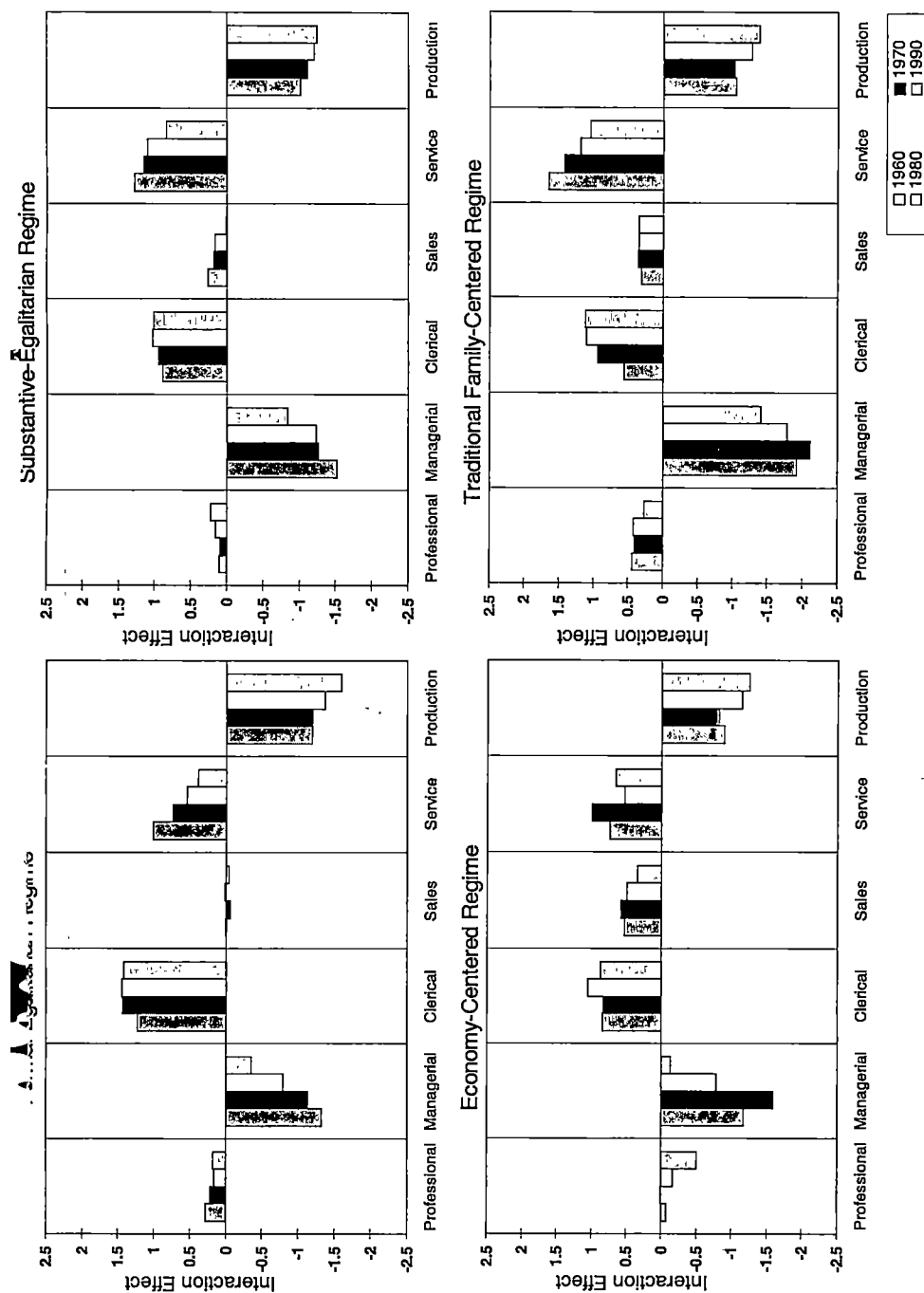


FIG. 4.—Trends in occupational sex segregation within regimes: 1960, 1970, 1980, and 1990

no authority, and the presence of women in these jobs does not necessarily indicate that women have achieved positions of power or prestige.

The dominant characteristic of the traditional family-centered regime is the continued high (although declining) levels of segregation in the managerial occupations, no doubt because managerial work is seen as interfering with women's family responsibilities or because it is otherwise perceived as less suitable for women. This result is consistent with expectations.

It is also informative to compare the average scale values across regimes for the four time periods. Figure 5 highlights the distinct trajectories of change across types. Among these trends, most striking is the desegregation of the managerial category in three of the regimes, all but the traditional family-centered. The high levels of segregation in the managerial category found in the substantive-egalitarian regime in 1990 was surprising.²⁹ This result may be partly attributable to the large percentage of women in these countries who choose to work part-time, an option that is not usually available in high-level administrative and managerial occupations. Statistical discrimination against women may also be a contributing factor since employers may be reluctant to hire women for such positions knowing that women are more likely than men to choose to make use of generous family-friendly benefits. It is also possible that the comparably high levels of segregation in the managerial category may be a by-product of women's movement into service occupations, which would partially offset their gains in the managerial occupations. The production category is the most integrated in the economy-centered and the traditional family-centered regimes. In this respect, one might say that women are being integrated into the "bottom" of the occupational structure in these regimes.

Levels of segregation in the service category for the traditional family-centered regime are almost as high as levels in the substantive-egalitarian regime. While the comparatively high segregation in the service category in the substantive-egalitarian regime is most likely due to the exportation of many of the functions of the family to the state, the overrepresentation of women in the service category in the traditional family-centered regime most likely arises because many of the service occupations are viewed as natural "extensions" of women's roles within the family and household. In this sense, quite diverse "logics" of gender segregation can sometimes generate similar results. Indeed, such a result highlights the necessity of examining patterns of occupational sex segregation *within* a broader soci-

²⁹ Wright, Baxter, and Birkelund (1995) show that despite the gender-egalitarian culture of the Scandinavian countries, Sweden and Norway lag behind the United States in terms of women's access to authority in the workplace.

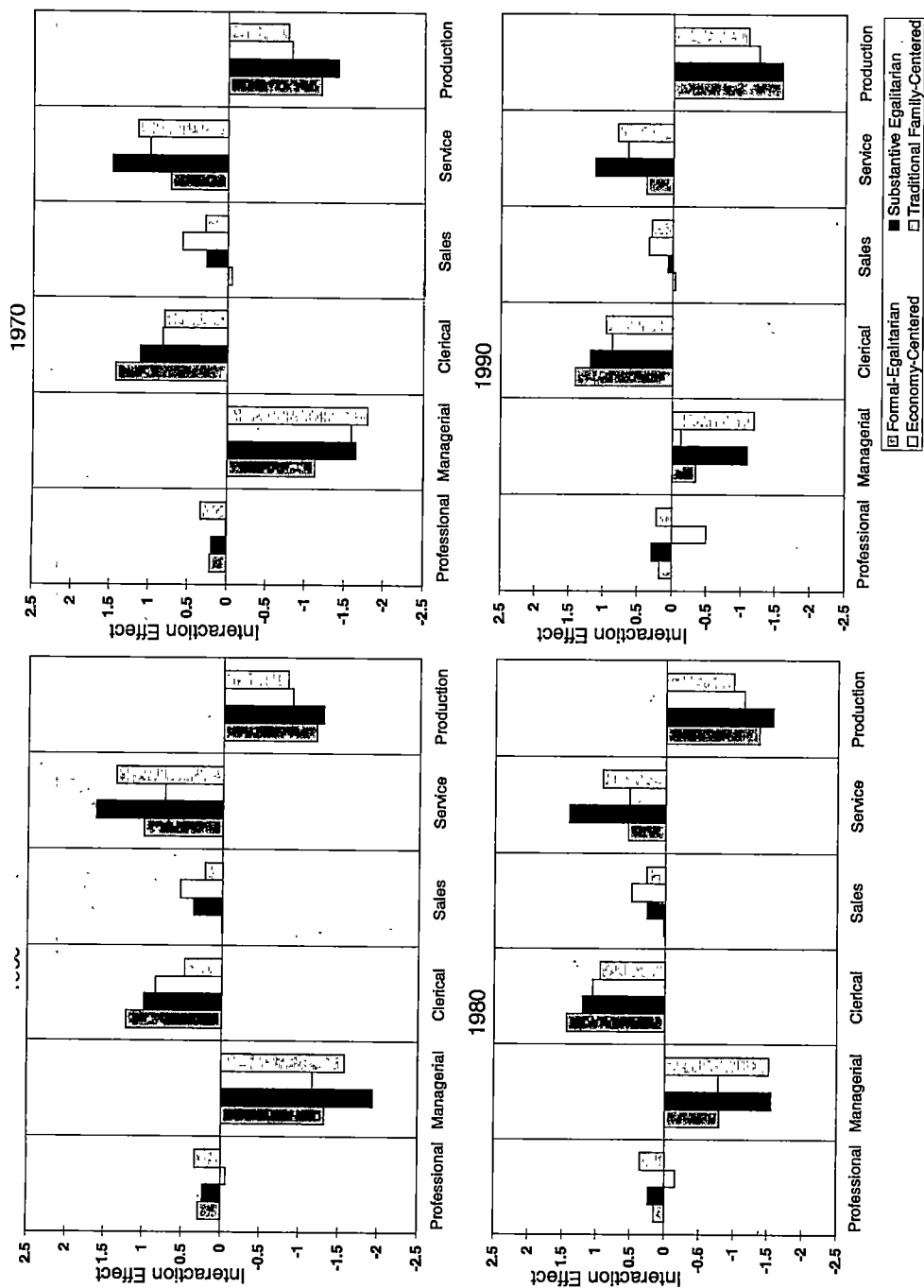


FIG. 5.—Profile of occupational sex segregation by regime: 1960, 1970, 1980, and 1990

TABLE 5

VARIANCE OF THE OCCUPATION-SPECIFIC SCALE VALUES BY TYPE OF REGIME: 1960,
1970, 1980, AND 1990

Year	Professional	Managerial	Clerical	Sales	Service	Production	Average
Formal egalitarian:							
1960025	.008	.004	.017	.000	.024	.013
1970026	.001	.002	.007	.003	.016	.009
1980010	.010	.006	.002	.001	.005	.006
1990017	.001	.001	.003	.001	.012	.006
Substantive egalitarian:							
1960020	.040	.046	.062	.014	.010	.032
1970081	.035	.058	.022	.004	.009	.035
1980006	.090	.064	.015	.039	.029	.041
1990061	.162	.112	.011	.007	.072	.071
Traditional family centered:							
1960233	.274	.165	.174	.073	.154	.179
1970105	.232	.160	.105	.061	.164	.138
1980059	.078	.155	.085	.089	.169	.106
1990049	.071	.092	.074	.083	.212	.097

etal context of gender stratification. Clearly, the overrepresentation of women in the service occupations in these two regimes have different implications when examined within the broader context of women's employment and gender stratification.

Figure 5 also suggests the possibility of a "parallel trend" as levels of segregation increase in clerical and production categories and decrease in the service category across all four regimes. However, it is notable that the general ranking of each regime with respect to levels of segregation within the occupational categories remains fairly consistent from 1960 to 1990, suggesting that even if there is a convergent trend within some categories, each regime maintains distinct characteristics.

The results presented in table 3 and in figures 4 and 5 imply that distinct segregation regimes have evolved between 1960 and 1990. I shall next address whether there is also convergence *within* the proposed regime types themselves. If there is indeed convergence within regimes, one would expect the variance in occupation-specific scale values within each regime to decrease over time. As shown in table 5, the variances within the formal-egalitarian and substantive-egalitarian regimes are exceedingly small, suggesting that convergence has already been effectively achieved within these two types. Although there is evidence of a slight continuing decline in the formal-egalitarian variances, I would not make too much of this result. Indeed, because convergence has already been achieved

across these two countries, there is little room for further convergence. The same comment holds for the substantive-egalitarian countries, but in this case there is more variation within the occupational categories over time and some suggestion of slight divergence between 1980 and 1990.³⁰ Nevertheless, these changes are rather modest given the small amount of variance present within this regime. The most striking evidence of convergence is found in the traditional family-centered regime. Initially this category had the highest variance in scale values, but the countries within this regime are clearly converging toward a common pattern of segregation.

Thus far, results indicate that while there may be some suggestion of a convergent decline in levels of segregation across the 14 countries included here, cross-national variation in segregation patterns do not support a conventional convergence trajectory. However, there is some evidence of convergence within the four types identified here, as suggested by the shrinking variance within regime types. Moreover, the fit statistics from the model of segregation regimes indicate that as a whole, the typology becomes increasingly powerful over time in explaining cross-national diversity at the aggregate level.

DISAGGREGATE ANALYSES

The preceding analyses were based on highly aggregated data, which mask possible changes in the distribution of men and women within the occupational structure. For instance, it is often the case that the professional category as a whole is less segregated than other occupational categories, yet segregation within the professional category may actually be quite high. In order to examine segregation trends in greater depth, I employed more detailed occupational categories in four countries (i.e., the United States, Sweden, Japan, and Hungary), each representing one of the four categories in the typology. It is important to emphasize that analyses of this sort put convergence to a strong test: if convergence is occurring across four countries with such diverse segregation regimes, it is likely that convergence is occurring more generally.

I will once again employ the model of constant cross-national variability in order to determine if patterns of change are consistent with a conver-

³⁰ The divergence in the managerial category arises because this category is integrating at a faster pace in Sweden. Denmark is responsible for the divergence within the production and professional categories as the distribution of female workers shifted away from production occupations and toward professional occupations. This shift may be explained by women's increased participation in higher education between 1980 and 1990 in Denmark.

TABLE 6

FIT STATISTICS FROM LOG-LINEAR MODELS OF OCCUPATIONAL SEX SEGREGATION IN THE UNITED STATES, SWEDEN, JAPAN, AND HUNGARY: 1960, 1970, 1980, AND 1990

Model	Sample Size	L^2	df	Δ	BIC
1. Constant cross-national variability ($O*S*C + O*S*T + O*C*T + S*C*T$)	141,993,894	723,437	477	1.7	714,483
2. Cross-national invariance ($O*C + S*C + O*S$):					
1960	27,089,752	1,062,746	159	5.6	1,060,024
1970	15,929,186	501,193	159	5.1	498,556
1980	37,495,412	1,014,416	159	4.8	1,011,643
1990	61,479,544	1,269,713	159	3.9	1,266,862

NOTE.— O = occupation; S = sex; C = country; T = time; R = regime type; Δ (the index of dissimilarity) represents the percentage of cases that are misclassified under the relevant model.

gence story. As shown in table 6, this model misclassifies less than 2% of the individuals. However, the likelihood and BIC test statistics suggest that the saturated model is still favored, and hence the extent of cross-national variability in patterns of segregation is evidently changing over time. Once again, the interpretation of these results is somewhat ambiguous but clearly leaves open the possibility of convergence. In order to test for convergence more explicitly, I will examine the results from the model of cross-national invariance and also trends in levels and patterns of segregation across these four countries.

Figure 6 presents trends in levels of segregation from 1960 to 1990. We can see that, in 1960, Sweden had the highest levels of segregation, indicating that women or men were overrepresented in the average occupation by a factor of 16.7. But, over time, levels of segregation decreased dramatically in Sweden and somewhat less dramatically in the United States and Japan. Although Hungary was the least segregated in 1960, there is subsequently very little change, and the Hungarian case is therefore altogether average in overall segregation 30 years later. These results provide some evidence of convergence in *levels* of segregation between 1960 and 1990.

This conclusion is further supported with the fit statistics from the model of cross-nationally invariant segregation (see table 6). In 1960, 5.6% of the cases were misclassified under this model, whereas in 1990 that percentage drops to roughly 3.9%, revealing that *patterns* of segregation are becoming more similar across countries. When disaggregated data are employed, both levels and patterns of segregation are converging. Once again, it is important to point out that while these results are based on

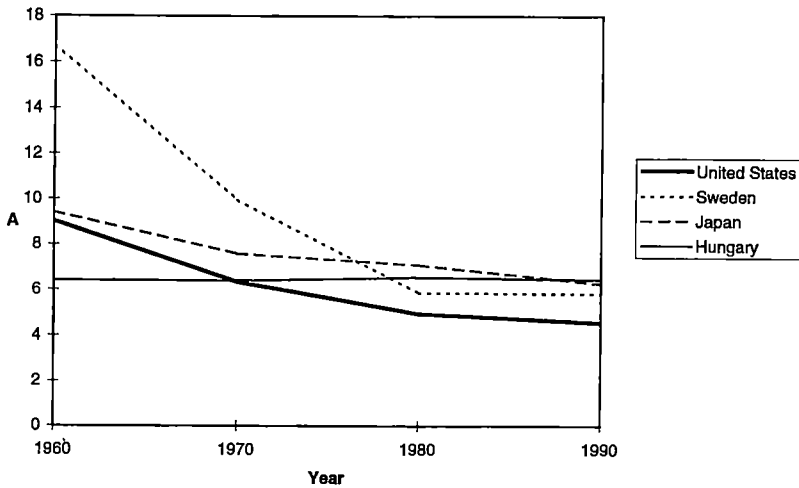


FIG. 6.—Trends in occupational sex segregation for the United States, Sweden, Japan, and Hungary, 1960–90.

only four countries, these countries were chosen primarily because they represent distinct segregation regimes. If convergence is found across these four diverse countries, then there is surely preliminary support for the argument that convergence may be occurring more broadly.

This evidence of convergence deserves further investigation, since the prior results with the aggregate data did not detect clear movement toward cross-national convergence. However, should further analysis show that such convergence is being driven by integrative changes *within* the major occupational categories and is therefore concealed by aggregate data, this seeming contradiction is resolved. I have proceeded, then, by first aggregating the data into the six-category classification and then re-computing occupation-specific measures of segregation. The resulting graphs, provided in figure 7, demonstrate that there is no evidence of convergence at the major occupational level (except perhaps in the clerical and service categories). This lack of convergence at the major category level parallels the results of the 14-country analyses presented previously.

However, figure 8 reveals that levels of segregation within the major occupational categories are indeed converging, with the possible exception of the unskilled category.³¹ For example, trends in the level of segregation within the clerical category show a remarkable convergence in levels of segregation over time, with Hungary in particular shedding its formerly discrepant position. It is worth noting that, within each major occupa-

³¹ The results for the managerial category should be interpreted with care since they were calculated from only two occupational groups.

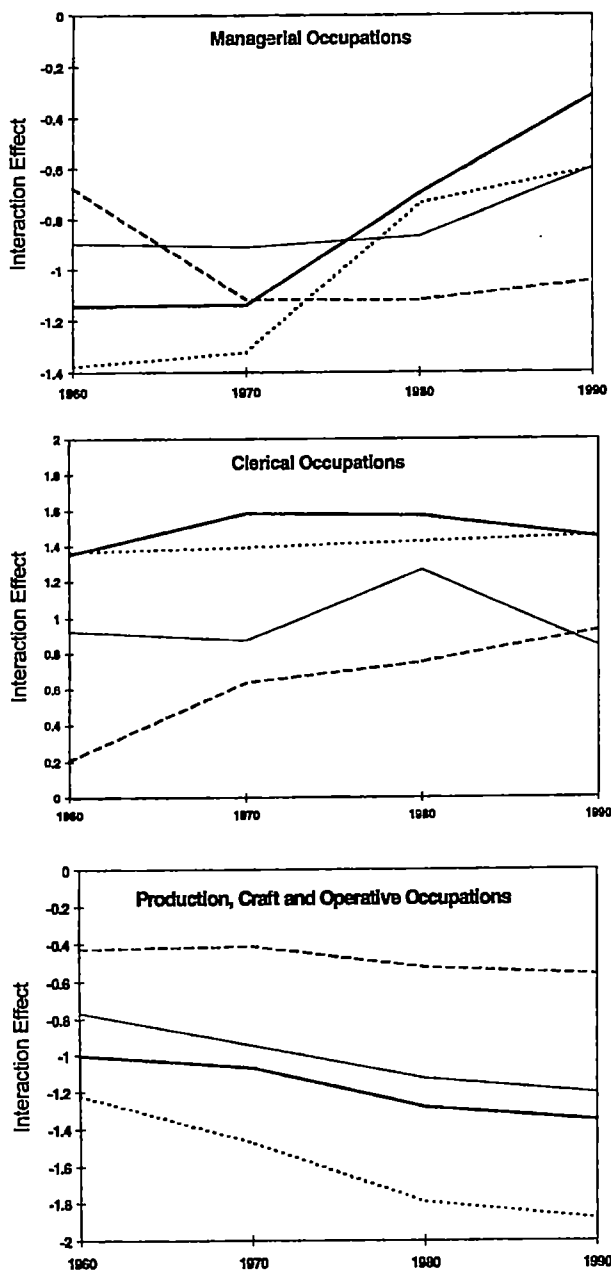


FIG. 7.—Trends in sex segregation across occupational categories, 1960–90. Dark unbroken line = United States; broken line = Japan; dotted line = Sweden; thin unbroken line = Hungary.

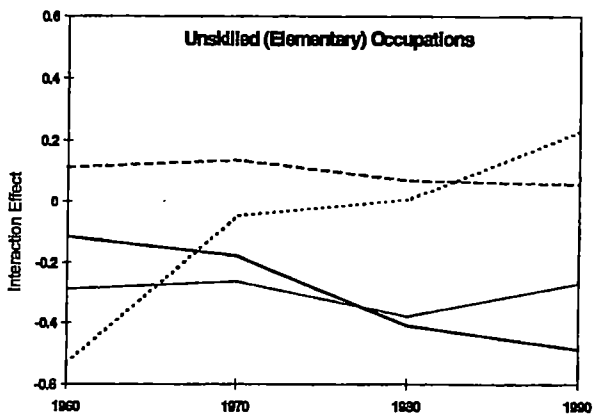
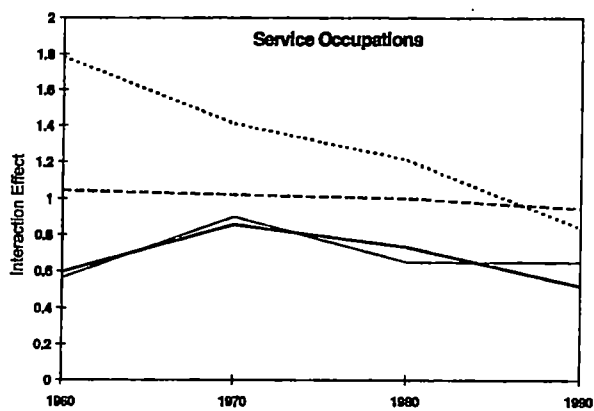
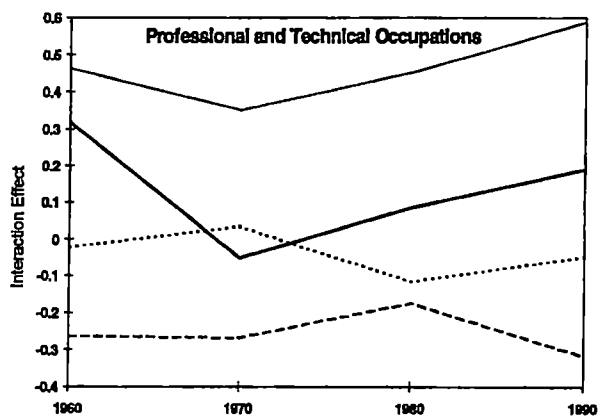


FIG. 7.—Continued

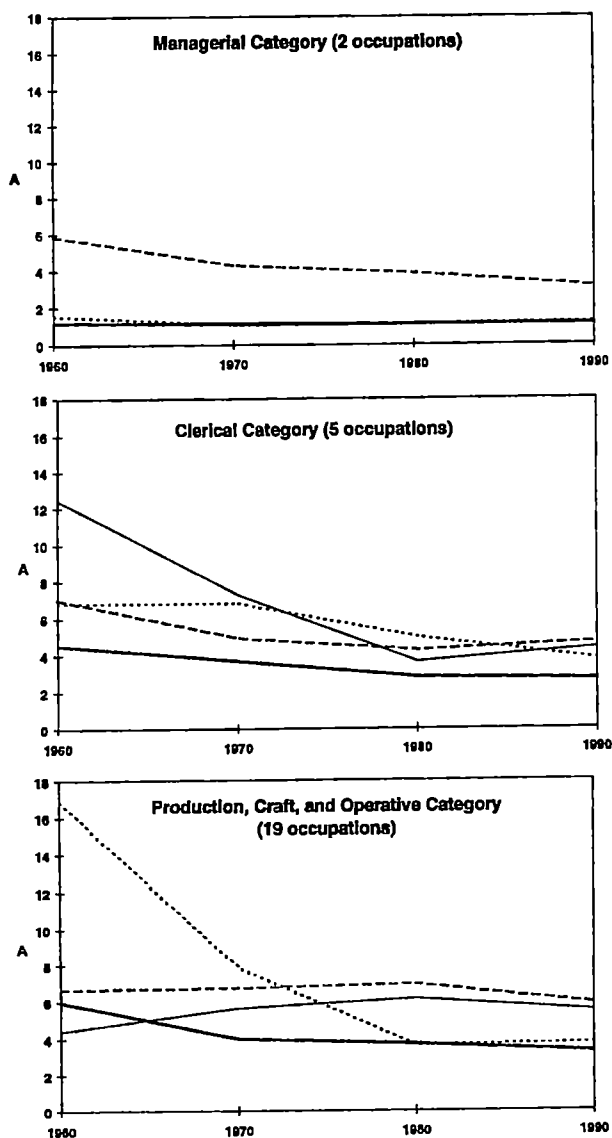


FIG. 8.—Trends in sex segregation within occupational categories, 1960–90. Dark unbroken line = U.S.; broken line = Japan; dotted line = Sweden; thin unbroken line = Hungary.

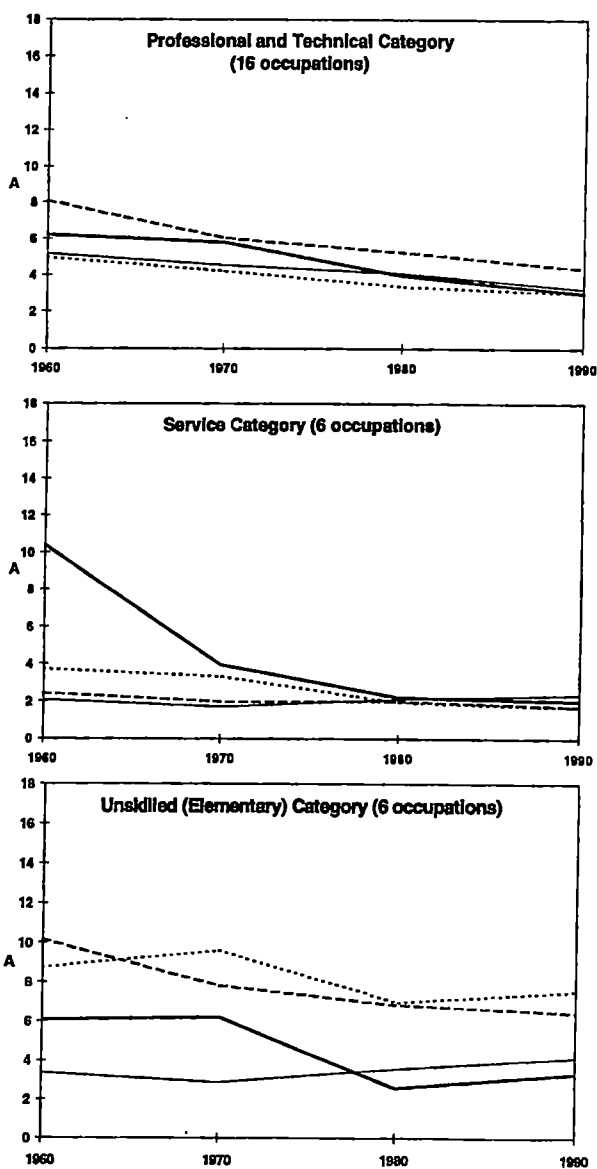


FIG. 8.—Continued

tional category, it is different countries that are making profound movements toward convergence, suggesting that such trends are not simply a function of a single outlier generated by occupational misclassification or dramatic social changes within one country.

DISCUSSION

The results presented here reveal that while there is some evidence of change in the extent of cross-national variation, such change as can be found at the aggregate (major category) level is not consistent with a conventional convergence story. However, this lack of convergence is not due to unstructured or idiosyncratic differences across countries but rather arises because the existing diversity in segregation is increasingly patterned according to one of the four segregation types. In this regard, there is partial convergence, not toward a single regime, but toward one of four possible regime types, each representing a different mixture of state response to gender inequality within the labor force. Furthermore, the amount of variability in occupation-specific values of segregation within some of the regime types (especially the traditional family-centered regime) is also declining over time, suggesting that within-type idiosyncrasies are withering away. The typology proposed here thus becomes increasingly powerful over time in accounting for cross-national diversity at the aggregate level.

When we turn to disaggregate analyses, we find that integration within the major occupational categories has resulted in cross-national convergence in the extent of segregation. In this respect, one might argue that a single regime of occupational sex segregation is emerging, with variation in both levels and patterns of segregation declining over time. However, this trend is driven by convergence *within* the major occupational categories and not *across* the major occupational categories. Hence, trends toward convergence in patterns of segregation are level specific; that is, results reveal evidence for continued diversity at the major-category level, and yet they also reveal convergence in segregation at the disaggregate level.

While these results may seem contradictory, they are actually complementary. Analyses at the aggregated and disaggregated level speak to different issues. The typology proposed here refers, in large part, to logics of gender distribution at the major occupational level. For example, substantive egalitarianism leads to the rise of a highly female-dominated service sector, while formal egalitarianism leads to a highly integrated managerial sector; but neither type precludes integration within the occupations that comprise these categories. At this more detailed level, egalitarian pressures of either the cultural or structural-functional variety are not easily countered, and one might accordingly expect trends toward integration.

In this regard, analyses at the major-category level reveal the macrolevel dynamics of organization by indicating the broad "types" of work in which men and women are engaged. A primary benefit of disaggregated tables, on the other hand, is that they capture more accurately the extent to which women and men are segregated in the more detailed aspects of the division of labor, thereby revealing the "horizontal" aspects of segregation.

It is worth asking whether these four segregation regimes are equally viable in the future. Given the growing attention focused on issues of gender equality, it is unlikely that countries can ignore altogether global and internal pressures to equalize opportunities for women to participate in the formal economy. The question that remains is *how* progress toward gender equality will be defined and brought about. Formal egalitarianism and substantive egalitarianism are both viable responses to issues of gender equality within the paid labor force. While they differ in their approach and conception of how to achieve "equality," the resulting variation is well within the normatively acceptable framework of gender equality.

The growing pervasiveness of gender egalitarian ideals implies that the traditional family-centered regime is the least likely to remain viable in the long run (although it will likely take much time for change to occur).³² Pressure to improve women's access to paid employment from both inside and outside traditional family-centered countries should eventually weaken the institutional structures that currently restrict women's access to full participation in the paid economy. While some women will undoubtedly choose to structure their labor force participation around family responsibilities, other women will most likely take advantage of the expanding opportunities (provided by equal opportunity and antidiscrimination legislation) to participate in the labor force continuously or to participate in male-dominated occupations. As a commitment to gender equality evolves, states may adopt either a substantive-egalitarian or formal-egalitarian framework. In all likelihood, the traditional family-centered countries will evolve toward formal egalitarianism, since this solution allows the state to respond to egalitarian pressures without directly challenging the fundamental assumption that child rearing properly lies within the domain of the family.

By contrast, most economy-centered regimes have substantive services for working mothers in place, and further pressures for gender equality may result in complementary labor market legislation. The subsequent

³² While it is true that countries that constitute the traditional family-centered regime are converging toward a common pattern of segregation (table 3), this does not preclude them from moving toward either substantive egalitarianism or formal egalitarianism in the future.

legislation may not be as extreme as that found in Sweden or the United States, but antidiscrimination and equal opportunity legislation should become increasingly common and enforceable. Over time, the economy-centered countries should therefore move toward a (perhaps weaker) form of substantive egalitarianism, with women overrepresented in the service occupations and somewhat less segregated within the managerial occupations. It is also possible that with time many substantive services will be abandoned, resulting in a more formal-egalitarian approach. For example, state-sponsored services in many Eastern European countries are currently losing support, at least in the short term, due to the growing influence of promarket forces.

In the coming decades, it is thus likely that countries will move among these different types as they wrestle with the often conflicting roles of women within the home and the labor force, as well as with economic fluctuations and changing cultural norms. While all features of the traditional family-centered and the economy-centered regimes will not disappear, the pressures toward gender egalitarianism will weaken these institutional solutions and force some movement toward alternative formal-egalitarian and substantive-egalitarian types.

APPENDIX A

Occupational Classification Scheme (54 categories; modified from ISCO-88 [ILO 1988])

Legislators, Senior Officials, and Managers (2 categories)

1. Legislators and senior officials
2. Corporate and general managers

Professionals, Technicians, and Associate Professionals (16 categories)

3. Physical, mathematical, engineering, computing, social science, and life science professionals, technicians, and associate professionals
4. Architects, engineers, and related professionals, technicians, and associate professionals
5. Medical doctors
6. Veterinarians
7. Dentists, pharmacists, and other health professionals (except nursing)
8. Nursing and midwifery professionals and associate professionals
9. College, university, and higher education teaching professionals
10. Secondary education teaching professionals

11. Primary and pre-primary education teaching professionals
12. Other teaching professionals
13. Business professionals, finance, sales, business services agents, customs, tax, and other associate professionals
14. Legal professionals
15. Other professionals (including archivists, librarians, writers, and creative or performing artists, religious professionals, and associate professionals)
16. Computer associate professionals and optical and electronic equipment operators
17. Modern health associate professionals (except nursing)
18. Professional, technical, and associate professionals, not elsewhere classified

Clerks (five categories)

19. Secretaries and keyboard operating clerks
20. Numerical clerks
21. Material-recording and transport clerks
22. Library, mail, and related clerks
23. Customer services clerks (cashiers, tellers, client information clerks)

Service Workers and Shop and Market Sales Workers (six categories)

24. Travel attendants and related workers
25. Cooks
26. Waiters, waitresses, and bartenders
27. Hair dressers, barbers, beauticians, and related workers
28. Other service (including housekeepers, personal care, protective service workers, etc.)
29. Models, salespersons, and demonstrators

Craft, Plant, and Machine Operators and Related Workers (19 categories)

30. Miners, shotfirers, stone cutters, and carvers, miscellaneous mining laborers
31. Building frame and related trades workers
32. Building finishers and related trades workers
33. Painters, building structure cleaners, and related trades workers
34. Metal moulders, welders, sheet-metal workers, structural-metal preparers, and related trades workers
35. Blacksmiths, tool-makers, and related trades workers
36. Machinery, electrical and electronic equipment mechanics, fitters, and assemblers
37. Precision workers in metal and related materials, potters, glass-makers, and related trades workers

38. Printing and related trades workers and machine operators
39. Food processing and related trades workers and machine operators
40. Wood treaters, cabinet-makers, and related trades workers and machine operators
41. Textile, garment, pelt, leather, and related trades workers and machine operators
42. Craftsmen and kindred workers, not elsewhere classified
43. Metal-processing-plant operators, glass, ceramics, and related plant operators
44. Chemical-processing-plant operators
45. Other machine operators and assemblers
46. Locomotive-engine drivers, ships' deck crews, and related workers
47. Motor-vehicle drivers
48. Agricultural and other mobile-plant operators

Unskilled ("Elementary") Occupations (six categories)

49. Domestic and related helpers, cleaners, and launderers
50. Building caretakers, window, and related cleaners
51. Messengers, porters, doorkeepers, and related workers
52. Construction labourers
53. Manufacturing labourers
54. Transport labourers and freight handlers

REFERENCES

- Anker, Richard. 1998. *Gender and Jobs*. Geneva: ILO.
- Bielby, William T., and James N. Baron. 1984. "A Woman's Place Is with Other Women: Sex Segregation within Organizations." Pp. 27-55 in *Sex Segregation in the Workplace: Trends, Explanations, Remedies*, edited by Barbara Reskin. Washington, D.C.: National Academy Press.
- Bradley, Karen. 1994. *The Incorporation of Women into the World's Systems of Higher Education, 1950-85*. Ph.D. dissertation. Stanford University, Department of Sociology.
- Brinton, Mary C. 1993. *Women and the Economic Miracle: Gender and Work in Post-war Japan*. Berkeley and Los Angeles: University of California Press.
- Brinton, Mary C., and Hang-Uue Ngo. 1993. "Age and Sex in the Occupational Structure: A United States-Japan Comparison." *Sociological Forum* 8:93-111.
- Castles, Francis G., ed. 1993. *Families of Nations*. Brookfield, Vt.: Dartmouth.
- Chang, Mariko Lin, Maria Charles, and Joon Han. In press. "Continuity and Change in Japanese Occupational Sex Segregation, 1950-1990." In *Comparative Perspectives on Sex Segregation*, edited by Maria Charles and David Grusky. Stanford, Calif.: Stanford University Press.
- Charles, Maria. 1992. "Cross-National Variation in Occupational Sex Segregation." *American Sociological Review* 57:483-502.
- . 1998. "Structure, Culture, and Sex Segregation in Europe." *Research in Social Stratification and Mobility* 16:89-116.

- Charles, Maria, and David Grusky. 1995. "Models for Describing the Underlying Structure of Sex Segregation." *American Sociological Review* 57:483-502.
- Christensen, Andrea. 1988. "Sex Discrimination and the Law." Pp. 329-47 in *Women Working: Theories and Facts in Perspective*, edited by Ann Helton Stromberg and Shirley Harkess. Palo Alto, Calif.: Mayfield Publishing.
- Corrin, Chris. 1994. *Magyar Women: Hungarian Women's Lives, 1960s-1990s*. New York: St. Martin's Press.
- Erikson, Robert, and John H. Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press.
- Esping-Andersen, Gøsta. 1990. *The Three Worlds of Welfare Capitalism*. Princeton, N.J.: Princeton University Press.
- . ed. 1993. *Changing Classes: Stratification and Mobility in Post-Industrial Societies*. Newbury Park, Calif.: Sage.
- . 1999. *Social Foundations of Postindustrial Economies*. New York: Oxford University Press.
- Fagan, Colette, and Jill Rubery. 1996a. "Transitions between Family Formation and Paid Employment." Pp. 348-78 in *International Handbook of Labour Market Policy and Evaluation*, edited by Günther Schmid, Jacqueline O'Reilly, and Klaus Schömann. Brookfield, Vt.: Edward Elgar.
- . 1996b. "The Salience of the Part-Time Divide in the European Union." *European Sociological Review* 12:227-30.
- Funk, Nanette, and Magda Mueller. 1993. *Gender Politics and Post-Communism*. New York: Routledge.
- Gardiner, Frances, ed. 1997. *Sex Equality Policy in Western Europe*. New York: Routledge.
- Gornick, Janet C., Marcia K. Meyers, and Katherin E. Ross. 1997. "Supporting the Employment of Mothers: Policy Variation across Fourteen Welfare States." *Journal of European Social Policy* 7:45-70.
- Grusky, David B., and Maria Charles. 1998. "The Past, Present, and Future of Sex Segregation Methodology." *Demography* 35:497-504.
- Grusky, David B., and Robert M. Hauser. 1984. "Comparative Social Mobility Revisited: Models of Convergence and Divergence in 16 Countries." *American Sociological Review* 49:19-38.
- Hagen, Elisabeth, and Jane Jenson. 1988. "Paradoxes and Promises: Work and Politics in the Postwar Years." Pp. 3-16 in *Feminization of the Labor Force: Paradoxes and Promises*, edited by Jane Jenson, Elisabeth Hagen, and Ceallaigh Reddy. New York: Oxford University Press.
- Hakim, Catherine. 1993. "Segregated and Integrated Occupations: A New Approach to Analysing Social Change." *European Sociological Review* 9:289-314.
- Hevener, Natalie Kaufman. 1983. *International Law and the Status of Women*. Boulder, Colo.: Westview.
- ILO (International Labour Office). Various years. *Year Book of Labour Statistics*. Geneva: ILO.
- . 1968. *International Standard Classification of Occupations, ISCO-68*. Geneva: ILO.
- . 1987. "Women Workers: Protection or Equality?" *Conditions of Work Digest*, vol. 4. Geneva: ILO.
- . 1988. *International Standard Classification of Occupations, ISCO-88*. Geneva: ILO.
- . 1990. *Year Book of Labour Statistics: Retrospective Edition on Population Censuses 1945-89*. Geneva: ILO.
- . 1992. *Ratifications by Convention and by Country*. Geneva: ILO.
- Jacobs, Jerry. 1993. "Theoretical and Measurement Issues in the Study of Sex Segregation in the Workplace: Research Note." *European Sociological Review* 9:325-30.

- Jacobs, Jerry, and Suet Lim. 1992 "Trends in Occupational and Industrial Sex Segregation in 56 Countries, 1960–1980." *Work and Occupations* 19:450–86.
- Kamerman, Sheila, and Alfred Kahn, eds. 1981. *Child Care, Family Benefits, and Working Parents*. New York: Columbia University Press.
- . 1991. *Child Care, Parental Leave, and the under 3s: Policy Innovation in Europe*. New York: Auburn House.
- Kerr, Clark, John Dunlop, Frederick Harbison, and Charles Myers. [1960] 1994. "Industrialism and Industrial Man." Pp. 659–69 in *Social Stratification: Class, Race and Gender in Sociological Perspective*, edited by David Grusky. Boulder, Colo.: Westview Press.
- Kolberg, Jon Eivind, ed. 1991. *The Welfare State as Employer*. New York: M. E. Sharpe.
- Ladin, Sharon. 1994. A Report on the Thirteenth Session of the Committee on the Elimination of Discrimination against Women. University of Minnesota, Humphrey Institute of Public Affairs, International Women's Rights Action Watch.
- Lewis, Jane. 1992. "Gender and the Development of Welfare Regimes" *Journal of European Social Policy* 2:159–73.
- Meyer, John W. 1994. "The Evolution of Modern Stratification Systems." Pp. 730–37 in *Social Stratification: Class, Race and Gender in Sociological Perspective*, edited by David Grusky. Boulder, Colo.: Westview Press.
- Moss, Peter. 1990. *Childcare in the European Community 1985–1990*. Brussels: Commission of the European Communities, Women of Europe Supplements.
- Nielsen, Ruth. 1983. *Equality Legislation in a Comparative Perspective*. Copenhagen: Women's Research Centre in Social Science.
- Norris, Pippa. 1987. *Politics and Sexual Equality: The Comparative Position of Women in Western Democracies*. Boulder, Colo.: Lynne Rienner Publishers.
- O'Connor, Julia S. 1993. "Gender, Class and Citizenship in the Comparative Analysis of Welfare State Regimes: Theoretical and Methodological Issues." *British Journal of Sociology* 44:501–18.
- OECD (Organisation for Economic Cooperation and Development). 1994. *Women and Structural Change: New Perspectives*. Paris: OECD.
- . 1998. *Employment Outlook*. Paris: OECD.
- Orloff, Ann. 1993. "Gender and the Social Rights of Citizenship: The Comparative Analysis of State Policies and Gender Relations." *American Sociological Review* 58:303–29.
- Parsons, Talcott. [1970] 1994. "Equality and Inequality in Modern Society, or Stratification Revisited." Pp. 670–85 in *Social Stratification: Class, Race and Gender in Sociological Perspective*, edited by David Grusky. Boulder, Colo.: Westview Press.
- Phillips, Angela, and Peter Moss. 1989. *Who Cares for Europe's Children?* Luxembourg: Office for Official Publications of the European Communities.
- Quadagno, Jill. 1988. "Women's Access to Pensions and the Structure of Eligibility Rules: Systems of Production and Reproduction." *Sociological Quarterly* 29:541–58.
- Ramirez, Francisco O., and Yun-Kyung Cha. 1990. "Citizenship and Gender: Western Educational Developments in Comparative Perspective." *Research in Sociology of Education and Socialization* 9:153–73.
- Ramirez, Francisco O., and Jane A. Weiss. 1979. "The Political Incorporation of Women." Pp. 238–49 in *National Development and the World System: Educational, Economic and Political Change 1950–1970*, edited by John Meyer and Michael Hannan. Chicago: University of Chicago Press.
- Reskin, Barbara. 1993. "Sex Segregation in the Workplace." *Annual Review of Sociology* 19:241–70.
- Reskin, Barbara, and Patricia Roos. 1990. *Job Queues, Gender Queues: Explaining Women's Inroads into Male Occupations*. Philadelphia: Temple University Press.

- Rubery, Jill, and Colette Fagan. 1995. "Gender Segregation in Societal Context." *Work, Employment and Society* 9:213–40.
- Rubery, Jill, Colette Fagan, and Friederike Maier. 1996. "Occupational Segregation, Discrimination and Equal Opportunity." Pp. 431–61 in *International Handbook of Labour Market Policy and Evaluation*, edited by Günther Schmid, Jacqueline O'Reilly, and Klaus Schömann. Brookfield, Vt.: Edward Elgar.
- Rueschemeyer, Marilyn, and Szonja Szelényi. 1989. "Socialist Transformation and Gender Inequality: Women in the GDR and in Hungary." Pp. 81–109 in *East Germany in Comparative Perspective*, edited by David Childs, Thomas A. Baylis, and Marilyn Rueschemeyer. London: Routledge.
- Ruggie, Mary. 1984. *The State and Working Women: A Comparative Study of Britain and Sweden*. Princeton, N.J.: Princeton University Press.
- Sainsbury, Diane, ed. 1995. *Gendering Welfare States*. Newbury Park, Calif.: Sage.
- . 1996. *Gender, Equality and Welfare States*. New York: Cambridge University Press.
- Schmidt, Manfred G. 1993. "Gendered Labour Force Participation." Pp. 179–237 in *Families of Nations*, edited by Francis G. Castles. Brookfield, Vt.: Dartmouth.
- Siaroff, Alan. 1994. "Work, Welfare and Gender Equality: A New Typology." Pp. 82–100 in *Gendering Welfare States*, edited by Diane Sainsbury. Newbury Park, Calif.: Sage.
- Siltanen, Janet, Jennifer Jarman, and Robert M. Blackburn. 1995. *Gender Inequality in the Labour Market*. Geneva: ILO.
- Steinberg, Ronnie, and Alice Cook. 1988. "Policies Affecting Women's Employment in Industrial Countries." Pp. 307–28 in *Women Working*, edited by Ann Helton Stromberg and Shirley Harkess. Palo Alto, Calif.: Mayfield Publishing.
- United Nations. 1995. *Human Development Report*. New York: United Nations Development Programme.
- . 1996. Meeting of States Parties to the Convention on the Elimination of All Forms of Discrimination against Women, 9th meeting. New York: United Nations Division for the Advancement of Women, DPCSD.
- U.S. Department of Health and Human Services. 1983. *Social Security Programs throughout the World*. Washington, D.C.: U.S. Department of Health and Human Services.
- U.S. Department of Health, Education, and Human Services. 1975; 1979. *Social Security Programs throughout the World*. Washington, D.C.: U.S. Department of Health, Education, and Human Services.
- Watts, Martin. 1993. "Explaining Trends in Occupational Segregation: Some Comments." *European Sociological Review* 9:315–19.
- . 1998. "Occupational Gender Segregation: Index Measurement and Econometric Modeling." *Demography* 35:489–96.
- Weeden, Kim A. 1998. "Revisiting Occupational Sex Segregation in the United States, 1910–1990: Results from a Log-Linear Approach." *Demography* 35:475–87.
- Whitehouse, Gillian. 1992. "Legislation and Labour Market Gender Inequality: An Analysis of OECD Countries." *Work, Employment and Society* 6:65–86.
- Wilensky, Harold L. 1983. "Political Legitimacy and Consensus: Missing Variables in the Assessment of Social Policy." Pp. 51–74 in *Evaluating the Welfare State*, edited by Shimon Spiro and Ephraim Yuchtman-Yaar. New York: Academic Press.
- . 1990. "Common Problems, Divergent Policies: An 18-Nation Study of Family Policy." *Public Affairs Report* 31:1–3.
- . 2000. "Rich Democracies: Political Economy and Performance." Manuscript. University of California, Berkeley.
- Wright, Erik Olin, Janeen Baxter, and Gunn Elisabeth Birkelund. 1995. "The Gender Gap in Workplace Authority: A Cross-National Study." *American Sociological Review* 60:407–35.

Multinomial Logit Latent-Class Regression Models: An Analysis of the Predictors of Gender-Role Attitudes among Japanese Women¹

Kazuo Yamaguchi
University of Chicago

This article describes the method and application of multinomial logit latent-class regression models in sociological research. This group of models is the categorical analogue of latent-variable regression for continuous latent variables such as those employed in LISREL or LISCOMP; therefore, though never yet employed in sociological research, the potential usefulness of these models in such research is very high. These models are regression extensions of log-linear latent-class models with group variables. An application focuses on predictors of three latent classes of gender-role attitudes among Japanese women. These classes are labeled "traditional gender-role supporters," "prowork gender-equality supporters," and "antiwork gender-equality supporters." The analysis illustrates the usefulness of the method by identifying the characteristics of each class compared with the others. Other possible applications of the models are also discussed.

INTRODUCTION

This article presents multinomial logit latent-class regression models and their applications. These models are regression extensions of log-linear latent-class models with group variables (Clogg and Goodman 1984, 1985; McCutcheon 1987; Hagenaars 1990) to be used with a more general class

¹ This article is dedicated to the memory of the late Clifford C. Clogg and his contributions to latent-class models. This study is partially supported by National Science Foundation grant SBR 9515341 to the author and by the Alfred P. Sloan Center on Families, Children, and Work at the University of Chicago. I wish to thank Shelby Haberman for clarifying for me the issues involved in comparing models with different numbers of latent classes. I also thank the 1995 SSM Committee in Japan for the use of their data and Diana Gillooly for her editorial assistance. Direct correspondence to Kazuo Yamaguchi, National Opinion Research Center, University of Chicago, 1155 East 60th Street, Chicago, Illinois 60637. E-mail: yama@cicero.spc.uchicago.edu

of data and covariates. They are similar to models described under such different names as "concomitant-variable latent-class models" (Dayton and Macready 1988), "linear logistic latent-class models" (Formann 1992), or "latent-variable regression [models] for multiple discrete outcomes" (Bandein-Roche et al. 1997). Therefore, the models are not novel, and the primary aim of this article is didactic. There are several reasons why a didactic article on this topic is worthwhile.

First, despite their usefulness, these models have not been applied in sociological research. Nor are substantive applications of latent-class models in general abundant as of yet. I believe one reason for this disregard is that they have not been used as regression models; therefore, once the capacity of the models described in this article is known, the usefulness of latent-class models in social science research will be recognized more fully. While this article analyzes gender-role attitudes for an illustrative application, previous applications of latent-class analysis focused on a variety of variables, including not only attitudinal variables, such as attitude about universalistic/particularistic values (Goodman 1974b; Clogg and Goodman 1985), attitude toward nontherapeutic abortion (Haberman 1979; McCutcheon 1996), and attitude about responsibilities of government (Hagenaars 1990), but also behavioral, occupational, and life-history variables such as social-mobility patterns (Clogg 1981; Formann 1992) job characteristics (Berkelund et al. 1966), patterns of intergenerational within-family exchange (Hogan, Eggbeen, and Clogg 1993), patterns of political campaign participation (McCutcheon 1987), and patterns of drug-use progression (Collins and Wugalter 1992). While none of those studies but Hogan et al.'s (1993) were based on regression models, all the analyses *could* be based on the regression models described in this article.

Despite the variety of possible applications, the multinomial logit latent-class regression model, or similar models with an alternative link function described below, has been little used in social science research: in fact, I cannot find a single application in the sociological literature. So a good illustrative application of the model and a clarification of the model's characteristics will be very useful. In the sociological literature, the approach that Hogan et al. (1993) took in their analysis of intergenerational exchange in American families is closest in intention, but they adopted a two-step approximation method that does not use full information. For each respondent's distinct observed response pattern for the set of response variables, we can obtain from a given latent-class model the set of conditional probabilities of that respondent's belonging to various latent classes. Hogan et al. (1993) assumed that each respondent belongs to the latent class having the largest conditional probability for her response pattern. Then they applied the multinomial logit model using this most

likely latent-class status of each respondent as the state of the dependent variable. However, this assignment of each respondent to the most likely latent class may seriously distort the characteristics of the latent classes because it tends to create a systematic discrepancy between the original proportions of the latent classes and their proportions obtained after the assignment. Usually, large latent classes tend to share even larger proportions after the assignment, and small latent classes share even smaller proportions, possibly resulting in such contradictory outcomes as no single respondent being assigned to the small latent classes despite their presence. In the multinomial logit latent-class model described in this article, we make no assumption in order to assign each respondent to a latent class. Instead, each respondent is treated as having a probability set of belonging to various latent classes depending on her response pattern, as in the standard latent-class model without covariates, and on the covariates of latent classes. Hence, no distortion occurs in the composition of the latent classes.

Second, this group of models deserves attention because it is the categorical analogue of latent-variable regression for continuous latent variables, such as those employed in LISREL (Jöreskog and Goldberger 1975) or LISCOMP (Muthén 1983), and its potential usefulness in sociological research is great. Latent-variable regression models, including both factor-analytic models and latent-class models, are *measurement models* that assume observed, or manifest, variables do not directly reflect the theoretical concept researchers wish to analyze but are instead indicators of the concept and that the latent variable derived from those indicators more adequately reflects the concept. Factor-analytic models and latent-class models differ in the assumed scale of the latent variable and its indicators, which is continuous for the former and discrete for the latter. More specific differences also exist. Factor-analytic models often impose orthogonality among multiple latent dimensions and are concerned mainly with the clustering of variables associated with each dimension. Accordingly, the method is unlikely to reveal a situation where people's response patterns for a set of variables bifurcate for certain values of another set of variables. For example, there is a dimension of traditionality/nontraditionality in gender-role attitudes, and a certain heterogeneity in response patterns may exist only among those with traditional (or nontraditional) gender-role attitudes. Since latent-class models are concerned with identifying distinct response patterns for a given set of response variables, or indicators, they are capable of characterizing heterogeneity of response patterns involving such partial differentiation, which the factor-analytic approach usually fails to do. Such an example based on an analysis of gender-role attitudes is given later. However, if the latent variable is continuous in

nature, latent-class models will be inefficient; that is, they tend to employ an unnecessarily large number of parameters to approximate a continuous state by a set of discrete states.

Third, although the link function that characterizes the dependence of latent-class membership on covariates can be specified in various ways (Dayton and Macready 1988), including continuation ratio logit, cumulative logit, and so on, the use of a multinomial logit function for the link function has certain advantages. In addition to its flexibility in contrasting both ordered and unordered latent states (Clogg 1988; McCutcheon 1996), it generates a straightforward extension of such log-linear latent-class models for contingency table analyses that were described by Goodman (1974a, 1974b), Haberman (1979), Clogg and Goodman (1984, 1985), McCutcheon (1987), Hagenaars (1990), and Heinen (1996). On the other hand, both the continuation logit and cumulative logit link functions assume a set of ordered states, and, therefore, their application is limited. Even for ordinary regression analyses with a polytomous dependent variable, the multinomial logit model is used most frequently, and the model introduced in this article differs from the conventional multinomial logit model only in the fact that instead of an observed polytomous variable, a latent-class variable with a set of its indicators is used as the dependent variable. Hence, the group of discrete latent-variable regression models that employs multinomial logit link function deserves special attention. The *multinomial logit latent-class regression model* described in this article refers to the latent-class regression model that employs this particular link function.

Fourth, although similar models have been introduced since Dayton and Macready's (1988) study, their methodological foundation in terms of parameter estimation has been established only quite recently. Since Dayton and Macready (1988) used the simplex algorithm and Formann (1992) used the EM algorithm for parameter estimation, neither incorporated estimation of the standard error of covariate effects on latent classes; therefore, hypothesis testing for the effects of covariates on latent-class membership has been limited. Only the most recent article by Bandeen-Roche et al. (1997), which employs the Newton-Raphson algorithm, incorporates estimation of the standard error. Bandeen-Roche et al. (1997), however, did not mention that the existing DNEWTON program (Haberman 1988) can be used to implement the Newton-Raphson method for the group of models that employ the multinomial logit link function. In this article, I use Haberman's (1988) DNEWTON program based on a modified Newton-Raphson algorithm for parameter estimation in an illustrative application of the multinomial logit latent-class regression model. (See Winship and Mare [1989], Mare and Winship [1991], Mare [1994],

and Yamaguchi and Kandel [1997] for the use of this program in the analysis of incompletely cross-classified frequency data.) While the DNEWTON program was written for contingency table analysis, it can in fact be used to estimate parameters for models of multinomial responses with latent classes whether the observations are for individual data or for cross-classified frequency data—although the model's chi-squared statistic cannot be used for a goodness-of-fit test if individual data are used, it can be used to compare among nested models regardless of the type of data. The capacity to use individual data implies that there is no restriction on the scale of covariates that predict the latent class: we can employ interval-scale variables as well as categorical variables—though the amount of computer's random access memory (RAM) available may temporarily limit the use of individual data if the design matrix becomes too large when individual data are used. In appendix A, below, I describe how this program can be used for different types of input data.

The fifth and last reason why a didactic article on this topic is useful is that since sociologists are familiar with standard log-linear latent-class models, an article that places more emphasis than the existing literature about related regression models on a description of the relationship between those log-linear latent-class models and the multinomial logit latent-class regression models will be beneficial. This is especially helpful because the use of the DNEWTON program is based on this relationship.

The situations where it is useful to employ the multinomial logit latent-class regression model can be described as follows. We have a set of substantively related categorical response variables that we believe reflect substantively meaningful *latent* individual characteristics. We also believe that the analysis will be more reliable and more parsimonious if we focus on the explanation for this latent variable rather than on separate explanations for each response variable. Unlike the factor-analytic latent-variable model, our model also assumes that the latent variable's states are discrete. In other words, we assume that *a small number of distinct patterns* exist within the responses to the given set of indicator variables. We identify the number of latent classes by a preliminary latent-class analysis of the set of response variables without including covariates. Finally, we assume a multinomial logit regression model using a specified number of latent classes as the polytomous states of the dependent variable with a set of covariates as predictors of latent-class membership. Since the latent-class variable is not observed, we need to include simultaneously in the model the set of response variables as the indicator variables of latent classes. In the basic model described below, we also assume that no covariate has a direct effect on any of the response variables. I describe, however, models that relax this assumption later.

The application presented in this article focuses on identifying predictors of gender-role attitudes among Japanese women. The Japanese population is known to have a greater proportion of people who agree with the statement "A man's place is at work, and a woman's place is at home" than is the case in Western and some other Asian countries (*Sorifu Fujinmondai Tantoushitsu* 1984; Hara and Hiwano 1990). While men are more likely than women to hold this traditional gender-role attitude, the attitude is not uncommon among women in Japan. Indeed, the existence of traditional gender-role attitudes among women is often believed to be as strong a barrier as their existence among men to the attainment of gender equality of opportunity in education, employment, and occupation.²

This article focuses on five variables related to gender-role attitudes collected in the 1995 Social Stratification and Mobility Survey. From the variables, we identify three latent classes among women: (1) "traditional gender-role supporters," (2) "prowork gender-equality supporters," who value women's work lives, and (3) "antiwork gender-equality supporters," who do not value women's work lives. We also identify the predictors to differentiate the first from the second and the third from the second latent classes.

MODELS AND METHODS

Models

I first describe a model using the expression for the log-linear model in order to show the correspondence between the multinomial logit latent-class regression model and the log-linear latent-class model. I first assume cross-classified frequency data. Then I present a more general expression of the model for individual data.

For simplicity of exposition, I assume three categorical response variables: R , S , and T ; three categorical covariates: A , B , and C ; and a latent-class variable: Y . An example of response variables, employed later in this article, is a set of variables about gender-role attitudes. Covariates are such variables as education, income, and birth cohort. The model assumes local independence among R , S , and T and the set of A , B , and C , given Y . This implies that we assume (1) that responses R , S , and T are mutually

² Brinton (1990) reports that in Japan mothers have greater university aspirations for sons than do fathers but that parents do not differ in university aspirations for daughters. As a result, mothers tend to have stronger gender-discriminatory attitudes about children's education than fathers do.

independent reflections of a latent-class state and (2) that the set of covariates A , B , and C affect the observed responses R , S , and T only *indirectly* through the former's effects on latent-class membership. A log-linear model that makes these assumptions can be expressed, for example, as:

$$\log(F_{rstyabc}^{RSTYABC}) = \tau_{abc}^{ABC} + \lambda_r^R + \lambda_s^S + \lambda_t^T + \lambda_y^Y + \lambda_{ry}^{RY} + \lambda_{sy}^{SY} + \lambda_{ty}^{TY} + \lambda_{ya}^{YA} + \lambda_{yb}^{YB} + \lambda_{yc}^{YC}, \quad (1)$$

where $F_{rstyabc}^{RSTYABC}$ is the frequency expected from the model, the lambda parameters satisfy the standard constraints of summing to zero across categories of each variable, and τ_{abc}^{ABC} is the set of parameters needed to make the interactions among covariates A , B , and C *marginally saturated*. We do not estimate τ_{abc}^{ABC} parameters as shown below. In equation (1), λ_{ry}^{RY} , λ_{sy}^{SY} and λ_{ty}^{TY} represent the association of each response variable with the latent-class variable and indicate that there is no *direct* association among response variables. On the other hand, although the model of equation (1) hypothesizes only the pairwise association of each covariate with Y , that is, λ_{ya}^{YA} , λ_{yb}^{YB} , and λ_{yc}^{YC} , the model can include three-factor or higher-order interaction effects that involve the latent-class variable Y and two or more covariates such as λ_{yab}^{YAB} , which indicates the interaction effect of A and B on Y when Y is specified as the dependent variable.

Let $F_{++++abc}^{RSTYABC} = \sum_r \sum_s \sum_t \sum_y F_{rstyabc}^{RSTYABC}$ then the joint conditional probability distribution of response and latent-class variables R , S , T , and Y , given the set of values of covariates A , B , and C , $P_{rsty|abc} \equiv F_{rstyabc}^{RSTYABC} / F_{++++abc}^{RSTYABC}$ is given for the model of equation (1) as

$$P_{rsty|abc} = \exp(\lambda_r^R + \lambda_s^S + \lambda_t^T + \lambda_y^Y + \lambda_{ry}^{RY} + \lambda_{sy}^{SY} + \lambda_{ty}^{TY} + \lambda_{ya}^{YA} + \lambda_{yb}^{YB} + \lambda_{yc}^{YC}) / \left[\sum_r \sum_s \sum_t \sum_y \exp(\lambda_r^R + \lambda_s^S + \lambda_t^T + \lambda_y^Y + \lambda_{ry}^{RY} + \lambda_{sy}^{SY} + \lambda_{ty}^{TY} + \lambda_{ya}^{YA} + \lambda_{yb}^{YB} + \lambda_{yc}^{YC}) \right] \quad (2)$$

and does not include τ_{abc}^{ABC} parameters. Let

$$\begin{aligned} \tau_y^Y &\equiv \log \left[\sum_r \sum_s \sum_t \exp(\lambda_r^R + \lambda_s^S + \lambda_t^T + \lambda_y^Y + \lambda_{ry}^{RY} + \lambda_{sy}^{SY} + \lambda_{ty}^{TY}) \right] \\ &= \lambda_y^Y + \log \left[\left\{ \sum_r \exp(\lambda_r^R + \lambda_{ry}^{RY}) \right\} \left\{ \sum_s \exp(\lambda_s^S + \lambda_{sy}^{SY}) \right\} \right. \\ &\quad \left. \times \left\{ \sum_t \exp(\lambda_t^T + \lambda_{ty}^{TY}) \right\} \right]; \end{aligned} \quad (3)$$

then we obtain

$$\begin{aligned}
 P_{rsty|abc} = & \left[\exp(\tau_y^Y + \lambda_{ya}^{YA} + \lambda_{yb}^{YB} + \lambda_{yc}^{YC}) / \right. \\
 & \left. \left\{ \sum_y \exp(\tau_y^Y + \lambda_{ya}^{YA} + \lambda_{yb}^{YB} + \lambda_{yc}^{YC}) \right\} \right] \\
 & \left[\exp(\lambda_r^R + \lambda_{ry}^{RY}) / \left\{ \sum_r \exp(\lambda_r^R + \lambda_{ry}^{RY}) \right\} \right] \quad (4a) \\
 & \left[\exp(\lambda_s^S + \lambda_{sy}^{SY}) / \left\{ \sum_s \exp(\lambda_s^S + \lambda_{sy}^{SY}) \right\} \right] \\
 & \times \left[\exp(\lambda_t^T + \lambda_{ty}^{TY}) / \left\{ \sum_t \exp(\lambda_t^T + \lambda_{ty}^{TY}) \right\} \right] \\
 & \equiv P_{y|abc} P_{r|y} P_{s|y} P_{t|y} \quad (4b)
 \end{aligned}$$

where $P_{y|abc}$, $P_{r|y}$, $P_{s|y}$, and $P_{t|y}$ are respectively the first, second, third, and fourth components of equation (4a). Although we do not have an observation that corresponds to $P_{rsty|abc}$, we have an observation that corresponds to $\sum_y P_{rsty|abc}$ for each sample person.

Equation (4b) indicates that the conditional probability $P_{rsty|abc}$ becomes the product of four conditional probabilities. One, $P_{y|abc}$, is the probability of being in latent class y , given the values a , b , and c for covariates A , B , and C , and the other three are the conditional probabilities of having, respectively, response r for R , response s for S , and response t for T , given latent class y . In the later example analyzing gender-role attitudes, equation (4b) will indicate that the conditional joint probability $P_{rsty|abc}$ of the set of gender-role indicator variables and a latent class of gender-role attitude can be expressed as the product of the conditional probabilities of having responses r , s , and t respectively for the gender-role indicator variables R , S , and T for a latent class y , and the conditional probability of being in latent class y for a given set of values a , b , and c respectively for covariates A , B , and C , such as education, income, and birth cohort. Equation (4a) also shows that each of the four conditional probabilities is expressed by the multinomial logit equation.

More generally, suppose $P_{rsty,i}$ is the conditional probability of having a state set $\{r, s, t, y\}$ for a given individual i with covariates $x_{i\cdot}$; then the model

$$P_{rsty,i} = P_{y|x_i} P_{r|y} P_{s|y} P_{t|y} \quad (5)$$

which assumes conditional independence among variables R , S , T , and X , given latent-class variable Y , can be modeled by specifying $P_{r|y}$, $P_{s|y}$, and $P_{t|y}$ as in equation (4a) and specifying $P_{y|xi}$ as follows:

$$P_{y|xi} = \exp\left(\tau_y^Y + \sum_k b_{yk} X_{ik}\right) / \left\{ \sum_y \exp(\tau_y^Y + \sum_k b_{yk} X_{ik}) \right\}, \quad (6)$$

where b_{yk} is the effect of variable X_k on the y th state of latent variable Y . Without loss of generality, we may set $b_{1k} = 0$ to make the first latent class the baseline state. This equation is equivalent to that of the standard multinomial regression equation with contrast-specific intercepts τ_y^Y and contrast-specific effects b_{yk} for covariates X_k , except that the dependent variable is not observed but latent. Note that intercept parameter τ_y^Y in equation (6) is the same as that defined in equation (3) and, therefore, becomes a nonlinear function of the lambda parameters that depend on variables R , S , T , and Y . In order for the model's parameters to be estimable by the DNEWTON program, we have to employ λ_y^Y instead of τ_y^Y , which leads to the following reparametrization for $P_{rsty,i}$:

$$P_{rsty,i} = \exp\left(\lambda_r^R + \lambda_s^S + \lambda_t^T + \lambda_y^Y + \lambda_{ry}^{RY} + \lambda_{sy}^{SY} + \lambda_{ty}^{TY} + \sum_k b_{yk} X_{ik}\right) / \left[\sum_r \sum_s \sum_t \sum_y \exp\left(\lambda_r^R + \lambda_s^S + \lambda_t^T + \lambda_y^Y + \lambda_{ry}^{RY} + \lambda_{sy}^{SY} + \lambda_{ty}^{TY} + \sum_k b_{yk} X_{ik}\right) \right]. \quad (7)$$

A comparison of this equation (7) with that of the corresponding log-linear model (equation [2]) shows that equation (7) is obtained by replacing $\lambda_{ya}^{YA} + \lambda_{yb}^{YB} + \lambda_{yc}^{YC}$ in equation (2) with $\sum_k b_{yk} X_{ik}$. This generalization implies two things. First, the multinomial regression equation is now expressed at the individual level i , rather than at the aggregate frequency level. Second, covariates x can now include continuous variables as well as categorical variables.

The likelihood function \mathcal{L} is given as

$$\mathcal{L} = \Pi_i \left[\sum_y \{P_{y|xi}(\Pi_r P_{r|y}^{\delta(r,i)})(\Pi_s P_{s|y}^{\delta(s,i)})(\Pi_t P_{t|y}^{\delta(t,i)})\} \right], \quad (8)$$

where Kronecker's dummy variable $\delta(r, i)$ takes the value "1" if the observed value of R is r for person i , that is, if $R_i = r$, and takes the value "0" otherwise. Variables $\delta(s, i)$ and $\delta(t, i)$ are similarly defined for S

and T . This likelihood function indicates that in the maximum likelihood estimation, parameters are estimated to maximize the product over sample individuals of the sum over latent classes of the product of four conditional probabilities, one of which indicates the dependence of latent-class membership on covariates and the others of which indicate the dependence of responses R , S , and T on latent-class membership. Hence, each sample person i is assumed to have a probability set of belonging to various latent classes, and the probabilities are determined by both her response pattern $\{r, s, t\}$ and her covariate values X_i , and, as shown in equation (7), the extent of association of particular latent classes with these two groups of variables are expressed in log odds ratios.

Although the model that uses parameter λ_y^V and the model that uses parameter τ_y^V obtained by using equation (3) are mathematically equivalent, these two models are regarded as having distinct dependent variables for the multinomial logit equation. The model that uses parameter τ_y^V includes multinomial regression equation (6) as a component, and since the dependent variable of this regression equation is the latent-class variable Y , we can call the model the multinomial logit latent-class regression model. On the other hand, the model of equation (7), which uses parameter λ_y^V , can be regarded as the multinomial logit model where the dependent variable has $N_R \times N_S \times N_T \times N_Y$ states, for N_R , N_S , N_T , and N_Y , which are the number of categories of variables R , S , T and Y respectively. This multinomial logit model specifies $(N_Y - 1) + N_Y(N_R + N_S + N_T - 3)$ special contrasts that correspond to $N_Y - 1$ parameters for factor λ_y^V and $N_Y(N_R + N_S + N_T - 3)$ parameters for the set of factors $(\lambda_r^R, \lambda_s^S, \lambda_t^T, \lambda_{ry}^{RV}, \lambda_{sy}^{SV}, \lambda_{ty}^{TV})$. Only the first set of $N_Y - 1$ parameters, which contrast each state against the baseline state of the latent-class variable, depend on covariates X in the regression equation; and the other $N_Y(N_R + N_S + N_T - 3)$ parameters have only the intercept for regression equation. In both of these parameterizations of the model, we can interpret the set of $N_Y - 1$ regression coefficients for each covariate X in terms of its effect on the log-odds of being in each of $N_Y - 1$ latent classes rather than the baseline latent class. Hence, the interpretation of the parameter estimates regarding the effects of covariates on latent-class membership is completely identical between the model with parameter λ_y and the model with parameter τ_y^V , and the expression using λ_y instead of τ_y^V is made only for the purpose of presenting the entire model as a log-linear latent-class model whose parameters are estimable by the DNEWTON program.

Although the model described above assumes that covariates X affect response variables only indirectly through their effects on the latent-class variable Y , this condition can be relaxed as follows. Generally, for log-linear latent-class models with group variables as covariates of the latent-class variable, we can hypothesize and test the direct effects of group vari-

ables on a particular response variable and its interaction with Y ; that is, we can make, for example, both λ_r^R and λ_{ry}^{RY} depend on group variables X (Clogg and Goodman 1984, 1985). We can hypothesize the same thing for the pairs $(\lambda_s^S, \lambda_{sy}^{SY})$ and $(\lambda_t^T, \lambda_{ty}^{TY})$ as well. These are models that hypothesize the response pattern of each latent class to depend on group variables. For the application of the multinomial logit latent-class regression models, however, it is undesirable to introduce three-factor interactions among a response variable, the latent variable Y , and a covariate X . The substantive meaning of latent classes is determined by the pattern of association of Y with the response variables. If we further allow this pattern of association to depend on covariates, the meaning of contrasts between being in one latent class versus another class changes with covariates, which is clearly undesirable in comparing regression coefficients across variables. Hence, even though a statistical improvement may be obtained by doing so, we should not introduce such three-factor interactions into the multinomial logit latent-class regression models; we make latent classes reflect the most statistically significant latent division of the entire population—even if such a division is not the most significant one among some subgroups of the population.

On the other hand, hypothesizing direct effects of covariates on a response variable without introducing the three-factor interaction involving the latent variable Y may not be a bad idea—though we lose as well as gain, as described below. For example, dependence on age may differ among the response variables. If we make only the latent-class variable depend on age, we may not be able to reflect such differences in age dependence among the response variables. However, if we introduce the direct effect of age, say X_1 , on R , the new effect of X_1 on Y is adjusted in such a way that the average of the effect of X_1 on R and the new effect of X_1 on Y corresponds to the old effect of X_1 on Y . Hence, the effect of X_1 on Y cannot be compared with the effects of other covariates on Y , and this caveat must be kept in mind in comparing the effects among covariates.

Methods of Parameter Estimation

Haberman's DNEWTON program is used for the maximum-likelihood estimation of parameters by a modified Newton-Raphson algorithm in the way described in appendix A. Although it is in principle possible to use individual data, the use of individual data makes the design matrix huge for models in the present application and turns out to require about 150 megabytes of random access memory (RAM), which I do not have. Because of this practical limitation, I employed cross-classified frequency data using categorical covariates so that the design matrix would be small enough to be run on a personal computer with 64 megabytes of RAM.

For the same reason, even when cross-classified frequency data are used, not all covariates are simultaneously included in a single model because the size of the design matrix depends on *the number of combinations of covariate states with nonzero observed frequency* and additional covariates increase the size of the design matrix. Hence, I employed nested models for the analysis by including/excluding some covariates in each model. In order to make bias due to the omission of covariates that have significant effects as small as possible, however, I tested models in such a way that I change the expression for the effects of covariates already included in the analysis into one that is more parsimonious, rather than omitting those covariates, before entering another set of covariates. The use of individual data as well as cross-classified frequency data with the DNEWTON problem is described in appendix A.

DATA AND VARIABLES

Data are taken from the 1995 Social Stratification and Mobility Survey in Japan. Women ages 25–64 at the survey date are included in the following analysis. Five response variables of gender-role attitudes indicate whether each respondent agrees or disagrees with the following five statements (my translation): “Q1: Men should work outside the home and women should maintain the home.” “Q2: Men and women should be brought up in different manners.” “Q3: Women are more suited for household work and child rearing than men are.” “Q4: Through their family roles, full-time housewives make valuable contributions to society.” “Q5: Work life should be an important element of life for women just like for men.” The possible responses in the survey are “agree,” “somewhat agree,” “somewhat disagree,” “disagree,” and “don’t know [DK]”. In the analysis, I created a dichotomous variable by combining “agree” with “somewhat agree” and “disagree” with “somewhat disagree” and then combining the DK answers to the majority. I added the DK answers to the majority because they constitute fractions too small to sustain separate categories (varying from 1.5% to 6% depending on the item). Their listwise omission nonetheless makes the sample less representative because missing responses do not overlap much and the omitted cases exceed 10% of the sample in sum. In a preliminary analysis, I confirmed that results from the latent-class model without covariates do not differ qualitatively between the employed assignment method and the case of listwise omission, so the assignment does not distort the relationships among the response variables. It is also worthwhile to note that the method itself does not require a dichotomization of response variables and can handle, in principle, any number of categories for a variable. However, the use of ordered categories, such as “agree,” “somewhat agree,” “somewhat disagree,” and “disagree,” is

TABLE 1
CROSS-CLASSIFIED FREQUENCY OF FIVE GENDER-ROLE ATTITUDES

	Q3:Y				Q3:N			
	Q4:Y		Q4:N		Q4:Y		Q4:N	
	Q5:Y	Q5:N	Q5:Y	Q5:N	Q5:Y	Q5:N	Q5:Y	Q5:N
Q1:Y:								
Q2:Y	118	51	36	12	11	3	6	4
Q2:N	92	33	30	13	24	11	6	7
Q1:N:								
Q2:Y	74	20	34	7	23	10	25	4
Q2:N	135	21	92	29	99	15	106	36

NOTE.—See the text for the definition of variables Q1 through Q5. Y = agree with the statement; N = disagree with the statement.

likely to generate more latent classes with “mixed” response patterns, which are relatively unstable because the latent variable’s states become more continuous in nature. Hence, by dichotomization, we are choosing to emphasize *more discreteness* in latent attitudes. This choice is based in part on an assumption that respondents’ choices between strong and weak expressions may depend more on their preferences of expression than on the strength/weakness of their attitudes, and, therefore, the distinctions between “agree” and “somewhat agree” and between “somewhat disagree” and “disagree” are not as important as the distinction between “somewhat agree” and “somewhat disagree.” However, it is also an analytical choice because we usually obtain fewer latent classes of relatively clear-cut characteristics by choosing fewer categories for response variables—though this choice leads to some information loss. Table 1 presents cross-classified frequencies of the five variables after they are dichotomized in this manner.

As predictors of latent classes reflected by those five response variables, variables that are related to (1) social stratum and employment status, (2) lifestyle, and (3) birth cohort are employed. The lifestyle variables were selected as a result of a preliminary analysis in search of additional predictors of the smallest latent class, labeled “antiwork gender-equality supporters,” because the variables related to social stratum and employment status did not clarify their characteristics well.

The variables related to social stratum and employment status include (1) father’s occupation, (2) respondent’s own education, (3) own present occupation, (4) own present individual income, (5) husband’s present occupation, (6) husband’s present income, (7) own present employment sta-

tus, and (8) subjective assessment of promotion opportunity at the present job. Mother's education and mother's work experience were considered but were omitted because their effects were nonsignificant. In addition, marital status was considered, but its effect was nonsignificant when birth cohort and either husband's income or husband's occupation (which controls for the distinction between women with and without spouses) are controlled for or when birth cohort and employment status (which distinguishes between those with and without jobs) are controlled for; it is, therefore, omitted from the final models.

The four lifestyle variables indicate whether each respondent feels the following statements apply to her (my translation): "LS1: I have hobbies/lifework outside of my work and family lives," "LS2: I prefer to make my present life pleasant rather than live frugally to save and make my later life better," "LS3: I try to make my associations with other people wide and extensive," "LS4: I am concerned with acquiring and/or displaying refined tastes and manners." The frequency distributions of the five response variables and their covariates are presented in table 2.

ANALYSIS

Latent-Class Models without Covariates

Table 3 presents the results of applying latent-class models using the five response variables and not including any covariate. The results show that according to the likelihood ratio chi-squared test neither the model with one latent class, that is, the independence model, nor the model with two latent classes fits the data and that models with three or four latent classes fit the data.

The procedure for choosing among models with different numbers of latent classes is not established yet because the identifiability conditions—which are required for the likelihood test, AIC or BIC, and are needed for the asymptotic chi-square distribution under the alternative hypothesis—do not hold under the null hypothesis, which specifies certain boundary conditions of parameters, for this kind of test.³ However, I conclude that the three-class model is better than the four-class model because of the stability described below. I have not found for the two-class model or the three-class model any other solution that yields a similar chi-square value but differs significantly in the pattern of association between latent classes and response variables. In this regard, these models have stable solutions. For the model with four latent classes, however, I found at least

³ I am indebted to Shelby Haberman for clarifying this and related points for me in a personal communication.

TABLE 2
FREQUENCY DISTRIBUTION OF VARIABLES

Variable	Frequency	Proportion
Response variables:		
1. Q1: "Men should work outside the home and women should maintain the home."		
Agree	457	38.5
Disagree	730	61.5
2. Q2: "Men and women should be brought up in different manners."		
Agree	438	36.9
Disagree	749	63.1
3. Q3: "Women are more suited for household work and child rearing than men are."		
Agree	797	67.1
Disagree	390	32.9
4. Q4: "Through family roles, full-time housewives make valuable contributions to society."		
Agree	740	62.3
Disagree	447	37.7
5. Q5: "Work life should be an important element of life for women just like for men."		
Agree	911	76.6
Disagree	276	23.3
Predictor variables:		
1. Father's occupation:		
Professional	70	5.9
Administrative/managerial	117	9.9
Clerical	123	10.4
Sales	110	9.3
Farm/fishery	291	24.5
Skilled manual	185	15.6
Semi- and nonskilled	162	13.6
DK or nonresponse	129	10.9
2. Education:		
Elementary (prewar)	39	3.3
Women's high (prewar)	14	1.2
Junior high (postwar)	233	19.6
Senior high (postwar)	668	56.3
Junior college (postwar)	137	11.5
College (postwar)	96	8.1
3. Present occupation:		
None	465	39.2
Professional	92	7.8
Clerical (administrative)	237 (9)	20.0
Sales	132	11.1
Farm/fishery	43	3.6
Skilled manual	76	6.4
Semi- and nonskilled	142	12.0
4. Individual income:		
None	379	31.9
Under 3.5 million yen	533	44.9
3.5 million yen and over	209	17.6
DK or nonresponse	66	5.6

TABLE 2 (Continued)

Variable	Frequency	Proportion
5. Present employment status:		
Nonemployed	465	39.2
Full-time employee	291	24.5
Part-time employee	234	19.7
Self-employed/family worker	197	16.6
6. Promotion opportunity:		
None	331	27.9
Some or much	856	72.1
7. Husband's present occupation:		
Professional	121	10.2
Administrative/managerial	81	6.8
Clerical	218	18.4
Sales	114	9.6
Farm/fishery	54	4.5
Skilled manual	198	16.7
Semi- and nonskilled	192	16.2
DK and nonresponse	40	3.4
Nonapplicable (no spouse)	169	14.2
8. Husband's present income:		
Under 4.5 million yen	345	29.1
4.5-7.5 million yen	316	26.6
7.5 million yen and over	192	16.2
DK and nonresponse	165	13.9
Nonapplicable (no spouse)	169	14.2
9. Birth cohort:		
25-34 in 1995	225	19.0
35-44 in 1995	326	27.5
45-54 in 1995	365	30.7
55-64 in 1995	271	22.8
11. LS1: "I have hobbies/lifework outside of my work and family lives."		
This applies to me	644	54.3
Other*	543	45.7
12. LS2: "I prefer to make my present life pleasant rather than live frugally to save and make my later life better."		
This applies to me	601	50.6
Other*	586	49.4
13. LS3: "I try to make my associations with other peo- ple wide and extensive."		
This applies to me	873	73.5
Other*	314	26.5
14. LS4: "I am concerned with acquiring and/or dis- playing refined tastes and manners."		
This applies to me	628	52.9
Other*	559	47.1

* This category includes both "This (statement) does not apply to me" and "I am not certain."

TABLE 3

TEST RESULTS OF LATENT-CLASS MODELS WITHOUT COVARIATES

	G^2	df	P
Independence	329.90	26	.000
Two latent classes	38.88	20	<.010
Three latent classes	20.30	14	>.050
Four latent classes	10.27	8	>.100

NOTE.— $N = 1,187$.

two other solutions whose chi-square values are very similar to that with the smallest chi-square value and yet differ substantially in their characteristics. In this regard, locally maximal solutions, nearly as good in attaining a fit with the data as the globally maximal solution, exist for the four-class model. It is therefore unlikely that the particular solution with the smallest chi-square value will remain the best model given small changes in the cross-classified frequency distribution of the response variables due to sampling variability. Hence, the solution for the four-class model is not stable. In the regression analysis presented in the next section, I assume the presence of three latent classes.

Table 4 presents estimates of (1) the proportion of each latent class and (2) the conditional probabilities of the two response-variable values for each given latent class, for the model with two latent classes, and for the model with three latent classes. Table 5 presents corresponding parameter estimates for these two models.

Although the results from table 5 indicate which response variables are significantly associated with which latent-class distinctions, equality between two latent classes of the conditional probabilities of responses for a given response variable can be tested more directly by testing models with an equality constraint. Appendix B, below, presents an additional analysis of tests for equality of response probabilities between latent classes for the three-class model. In particular, those results indicate that for the three-class solution, latent classes 1 and 3 do not differ significantly in their responses to Q2 and Q3, and all other differences between latent classes 1 and 2 and between latent classes 1 and 3 are significant. These results for equality constraints are consistent with the results of table 5.

From the substantive content of the response variables, we can expect that a person who supports gender differences in family and work roles, whom I call a *traditional gender-role supporter*, will agree with Q1, Q2, Q3, and Q4 and will disagree with Q5. On the contrary, a person who supports gender equality in family and work roles, whom I call a *gender-*

TABLE 4

THE PROPORTION AND CONDITIONAL PROBABILITIES OF RESPONSES OF EACH LATENT CLASS

	TWO-CLASS SOLUTION		THREE-CLASS SOLUTION		
	Class 1	Class 2	Class 1	Class 2	Class 3
Proportion51	.49	.40	.50	.10
Response probabilities:					
Q1:					
Y09	.69	.06	.69	.18
N91	.31	.94	.31	.82
Q2:					
Y17	.58	.17	.57	.14
N83	.42	.83	.43	.86
Q3:					
Y46	.90	.47	.89	.39
N54	.10	.53	.11	.61
Q4:					
Y47	.78	.50	.79	.31
N53	.22	.50	.21	.69
Q5:					
Y81	.73	1.00	.73	.00
N19	.27	.00	.27	1.00

equality supporter, will disagree with Q1, Q2, Q3, and Q4 and will agree with Q5. The results of table 4 show that for the solution of the model with two latent classes, the two latent classes are almost equal in size, and women in latent class 1 have the characteristics of gender-equality supporters, and those in latent class 2 have the characteristics of traditional gender-role supporters except for their pattern of responses to Q5. Women in latent class 1 have large probabilities of disagreeing with Q1 and Q2, less than 50% probabilities of agreeing with Q3 and Q4, with which more than 60% of women agree on average, and a large probability of agreeing with Q5. On the contrary, women in latent-class 2 have respectively 69% and 58% probabilities of agreeing with Q1 and Q2, with which less than 40% of women on average agree, and 90% and 78% probabilities of agreeing with Q3 and Q4, with which about two-thirds of women on average agree; regarding these response patterns, they are traditional gender-role supporters. However, women in latent-class 2 also tend to agree with Q5, which makes their attitudes on Q5 not greatly different from women in latent class 1. Results in table 5 show that the association between distinction of the two latent classes and distinction of two responses are strongly significant (at least the 0.1% level) for Q1, Q2, Q3, and Q4, but that association attains only the 5% level of significance for Q5.

TABLE 5
PARAMETER ESTIMATES FROM THE TWO MODELS

	Two-Class Solution	Three-Class Solution
λ^{Q1}	-1.163 (.220)	-1.377 (.324)
λ^{Q2}	-.794 (.091)	-.783 (.092)
λ^{Q3}	-.089 (.065)	-.061 (.067)
λ^{Q4}	-.061 (.059)	-.003 (.061)
λ^{Q5}709 (.062)	5.655 (.061)
λ_1^X153 (.546)	5.328 (.628)
λ_2^X	-1.031 (.453)
λ_1^{X*Q1}	1.570 (.216)	1.773 (.306)
λ_1^{X*Q2}948 (.104)	.933 (.107)
λ_1^{X*Q3}	1.166 (.157)	1.110 (.651)
λ_1^{X*Q4}705 (.090)	.652 (.091)
λ_1^{X*Q5}	-.216 (.092)	$-\infty^*$ (∞)
λ_2^{X*Q1}623 (.308)
λ_2^{X*Q2}	-.129 (.224)
λ_2^{X*Q3}	-.170 (.133)
λ_2^{X*Q4}	-.391 (.156)
λ_2^{X*Q5}	$-\infty^*$ (∞)

NOTE.—The absolute value of the standard error is given in parentheses. Variables Q1, Q2, Q3, Q4, and Q5 are coded 1 and -1 for the first and the second category, respectively (i.e., the deviation contrast is used). On the other hand, the dummy variable expression is used for latent classes using the first class as the baseline category.

* Although the significance of this coefficient is not testable by the z-test, the likelihood ratio chi-squared test comparing this model with the model that imposes 0 for this coefficient is significant at the .001 level.

The results of table 5 show for the model with two latent classes the extent of association (measured by the absolute value of the log odds ratio) between responses of each question and latent classes. The test of their pairwise differences reveals an order of Q1, Q3, Q2, Q4, and Q5, although not all differences are significant, as the order from the most to the least important predictors of the distinction between gender-equality supporters and traditional gender-role supporters.⁴ Thus question Q1 about whether “men should work outside the home and women should maintain the home” has the greatest predictive power, and Q5 about whether “work

⁴ $|\lambda^{X*Q1}|$ is significantly greater than $|\lambda^{X*Q2}|$, $|\lambda^{X*Q4}|$, and $|\lambda^{X*Q5}|$; $|\lambda^{X*Q3}|$ is significantly greater than $|\lambda^{X*Q4}|$ and $|\lambda^{X*Q5}|$; $|\lambda^{X*Q2}|$ is significantly greater than $|\lambda^{X*Q5}|$; and $|\lambda^{X*Q4}|$ is significantly greater than $|\lambda^{X*Q5}|$. Hence, even though only a partial order exists for pairwise comparisons, the set of greater and smaller lambda values differ among the questions and reveal the order of Q1, Q3, Q2, Q4, and Q5.

life should be an important element of life for women just like for men" has the least explanatory power. Q5, however, becomes the most important predictor of a subdivision of gender-equality supporters in the results from the model with three latent classes, as shown below.

The results from the model with three latent classes in table 4 show that the three-class solution is obtained by dividing the latent group of gender-equality supporters into two latent subgroups, while the other group, namely, traditional gender-role supporters, remains an undivided single latent group. The division in the first group occurs mainly because a minority of gender-equality supporters (10% of women), latent-class 3, almost completely disagree with the statement "Work life should be an important element of life for women just like for men," whereas a majority of gender-equality supporters (40% of women), latent-class 1, almost completely agree with the statement. Although the log odds ratios between variable Q5 and the contrasts of latent-class 3 versus 1 and latent-class 2 versus 1 both approach minus infinity as shown in table 5, and, therefore, the Z-test of these parameter estimates cannot be made—though the effects become significant at the 0.1% level in models with covariates as shown later. The results in table 5 also indicate that compared with women in latent-class 1, women in latent-class 3 are also somewhat less likely to disagree with Q1 and somewhat more likely to disagree with Q4. Hence, despite the fact that women in latent-class 3 do not consider work life to be as important an element of life for women as for men, they tend to disagree most strongly with the statement that full-time housewives make valuable contributions to society. This puzzling double denial of the value of both the work role and family role of women makes this group quite different from the other two latent groups.

Below, I analyze gender-role attitudes by focusing on the predictors of the three latent classes. I label latent-classes 1, 2, and 3, respectively, "pro-work gender-equality supporters," "traditional gender-role supporters," and "antiwork gender-equality supporters." The choice of the three-class model rather than the two-class model is based on a substantive concern with identifying the predictors of differentiation among gender-equality supporters regarding the evaluation of women's work lives, as much as the predictors of differentiation between traditional gender role supporters and gender-equality supporters.

Multinomial Logit Latent-Class Regression Models

It is worthwhile to make a caveat regarding the interpretation of results from multinomial logit latent-class regression models described below. Since this is an analysis of cross-sectional data, many covariates other than those of family background and education variables do not always

precede temporally the realization of observed response variable values. More generally, the causality of relations between response variables and covariates cannot be established by this study. Hence, we will be concerned only with the significance of the unique predictive power of covariates without making any causal arguments.

Table 6 presents the results from three models. Model 1 includes birth cohort and father's occupation as predictors of latent class, model 2 adds to model 1 respondent's own education as another predictor and model 3 is a simplification of model 2. Chi-squared statistics are not shown for these three models and for models whose results are presented in the following two tables because the models are based on different cross-classified aggregations of data in order to minimize the size of the design matrix for each model and, therefore, their chi-squared statistics are not comparable. Generally speaking, we can use a difference in chi-squared statistics between two nested models for a chi-squared test only if we use either the same set of cross-classified frequency data or the same individual data for the two models. When the parameter estimation for model 3 is based on the same set of cross-classified frequency data used for model 2, their chi-square difference becomes 9.2 with 14 degrees of freedom, which is not significant. Hence, model 3 is more parsimonious than model 2. Below, we are concerned only with the significance of regression coefficients and their implications.

Regarding the effects of birth cohort, the results of model 1 show that compared with people in younger birth cohorts, people in older birth cohorts are more likely to be traditional gender role supporters (latent-class 2) and less likely to be antiwork gender-equality supporters (latent-class 3) than prowork gender-equality supporters (latent-class 1). Although the effects of birth cohort can be confounded with age effects in the analysis of cross-sectional data, these trends indicate that if age effects are relatively small compared with unique birth-cohort effects, the proportions of people in both groups of gender-equality supporters have increased historically, but more so for the antiwork group in terms of the amount of increase in log odds, and the proportion of traditional gender-role supporters has historically decreased.

Regarding the effects of father's occupation, the results of model 1 show that compared with women having fathers in semiskilled and nonskilled manual occupations, women having fathers in professional and clerical occupations are less likely to be traditional gender-role supporters, and still less likely to be antiwork gender-equality supporters, than prowork gender-equality supporters. There are also weaker (at the 10% significance level) tendencies that compared with women having fathers in semiskilled and nonskilled manual occupations, women having fathers in sales or

farm/fishery occupations are less likely to be antiwork gender-equality supporters than prowork gender-equality supporters.

The results from model 2 (which adds to model 1 respondent's own education as a covariate) indicate that, controlling for birth cohort and father's occupation, compared with women having postwar high school education—where “postwar” implies post World War II—women having sex-segregated prewar education are more likely to be traditional gender role supporters than prowork gender-equality supporters and are not found at all among antiwork gender-equality supporters. There are no significant differences in latent-class membership among those having postwar junior high school, high school, and junior college education. Junior colleges in Japan are mostly women's junior colleges characterized by an emphasis on traditional women's majors such as home economics, education, and humanities. Compared with women having lower levels of postwar education, women who have four years of college education are significantly less likely to be traditional gender-role supporters or antiwork gender-equality supporters than prowork gender-equality supporters.

The results from model 3, which is a more parsimonious version of model 2, show the characteristics described above more concisely. I use this model as the baseline model for further additions of covariates.

Table 7 presents results from five additional models that add to model 3 some variables related to respondent's or her family's present position in the social stratification. In particular, model 4 adds respondent's present occupation, and model 5 adds respondent's husband's present occupation as a covariate.

While own occupation and husband's occupation are often regarded as being among the major indicators of social status for women, the results of models 4 and 5, rather surprisingly, are not very informative: there is no clear association between gender-role attitudes and occupational-status gradation. The most significant effect in the results of model 4 regarding the effects of own occupation is that if women have no occupation, that is, if they are not working, they are more likely than others to be traditional gender-role supporters. Compared with semiskilled and nonskilled manual workers, women having farm/fishery occupations are also more likely than others to be traditional gender-role supporters than prowork gender-equality supporters, and skilled workers are less likely than others to be antiwork gender-equality supporters than prowork gender-equality supporters.

As for the effects of husband's occupation, the only significant effect is that having a husband in a clerical occupation predicts a significantly smaller probability of being an antiwork gender-equality supporter. Although having no spouse also predicts smaller probabilities of being in

TABLE 6
MULTINOMIAL LOGIT LATENT-CLASS REGRESSION 1: THE EFFECTS OF FATHER'S OCCUPATION AND EDUCATION

CONTRAST OF LATENT CLASSES	MODEL 1			MODEL 2			MODEL 3		
	2 vs. 1	3 vs. 1		2 vs. 1	3 vs. 1		2 vs. 1	3 vs. 1	
I. Response variables: ^a									
Q1	2.167***	1.102***		2.006***	1.003***		1.982***	.972***	
Q2913***	-.178		.913***	-.191		.918***	-.118	
Q3	1.077***	-.124		1.124***	-.084		1.132***	-.087	
Q4629***	-.366***		.640***	-.333**		.642***	-.311**	
Q5	-2.461***	-5.471***		-1.895***	-4.596***		-1.937***	-4.621***	
II. Predictor variables:									
1. Birth cohort:									
Linear ^b354***	-.246*		.227**	-.302*		.258**	-.272*	
2.1. Father's occupation—I (vs. semi- and nonskilled):									
Professional	-.761*	-1.389*		-.434	-1.242*		
Administrative/managerial	-.222	-.481		-.010	-.419		
Clerical	-.786**	-1.742***		-.642***	-1.684***		
Sales	-.118	-.783†		-.024	-.756†		
Farm/fishery	-.033	-.675†		-.024	-.715*		
Skilled		-.434		.060	-.407		
DK and nonresponse	-.415	-.159		-.371	-.058		

2.2. Father's occupation—II (vs. blue collar and DK):

Professional/clerical	-.524**	-1.286***
Administrative/managerial	-.021	-.138
Sales/farm055	-.522*
3.1. Education (vs. postwar senior high):						
Prewar elementary	2.753**	-∞
Prewar high	1.511†	-∞
Postwar junior high134	.428
Postwar junior college	-.210	.505
Postwar college	-.759**	-.881†
3.2 Education (vs. postwar junior high, high, and junior college):						
Prewar	2.183***	-∞
Postwar college	-.733**	-.975*

^a Parameter estimates for the main effects of the response variables and the main effects of the latent-class variable are omitted. Response variables are coded "1" and "-1" for the first and second categories, and dummy variables are used for latent classes.

^b The linear cohort effect is obtained by giving the following scores for the four categories of birth cohort: 0 = 25-34 in 1995; 1 = 35-44 in 1995; 2 = 45-54 in 1995; 3 = 55-64 in 1995.

† $P < .10$.

* $P < .05$.

** $P < .01$.

*** $P < .001$.

TABLE 7
MULTINOMIAL LOGIT LATENT-CLASS REGRESSION — 2: THE EFFECTS OF OWN AND HUSBAND'S OCCUPATION AND INCOME

	MODEL 4			MODEL 5			MODEL 6			MODEL 7			MODEL 8		
	2 vs. 1	3 vs. 1	2 vs. 1	3 vs. 1	2 vs. 1	3 vs. 1	2 vs. 1	3 vs. 1	2 vs. 1	3 vs. 1	2 vs. 1	3 vs. 1	2 vs. 1	3 vs. 1	3 vs. 1
CONTRAST OF LATENT CLASSES															
I. Response variables (presentation omitted)															
II. Predictor variables:															
1. Birth cohort:															
Linear270***	-.220*	.224***	-.345**	.305***	-.236*	.248**	-.299**	.304***	-.261*					
2. Father's occupation—II (vs. blue collar and DK):															
Professional/clerical	-.528**	-1.287***	-.490*	-1.409***	-.586**	-1.333***	-.510**	-1.325***	-.556***	-1.303***					
Administrative/managerial	-.169	-.036	-.096	-.173	-.096	-.089	-.056	-.126	-.123	-.096					
Sales/farm022	-.688**	.084	-.518*	.039	-.538*	.082	-.511*	.066	-.534*					
3. Education (vs. postwar junior high, senior high, and junior college):															
Prewar	1.675**	—∞	1.978***	—∞	1.706***	—∞	1.802***	—∞	1.581**	—∞					
Postwar college	-.570*	-.879†	-.789**	-1.620**	-.458†	-.776†	-.694**	-1.011*	-.448†	-.762†					
4. Own present occupation (vs. semi- and nonskilled):															
Professional	-.429	-.545					
Clerical051	-.459					
Sales348	-.697					
Farm/fishery	1.139**	.864					
Skilled	-.244	-1.370*					
None	1.199***	-.033					

5. Husband's present occupation

(vs. semi- and nonskilled):

Professional
Administrative/managerial
Clerical
Sales
Farm/fishery
Skilled
DK and nonresponse
NA (no spouse)
6. Own income (vs. under 3.5 mil-

lion yen):

None
Over 3.5 million yen
DK and nonresponse

7. Husband's income (vs. 4.5-7.5

million yen):

Under 4.5 million yen
Over 7.5 million yen
DK and nonresponse
NA (no spouse)

† $P < .10$.

* $P < .05$.

** $P < .01$.

*** $P < .001$.

latent classes 2 and 3, these effects are spurious effects of having a smaller proportion of the nonemployed and a greater proportion of full-time employees, and they become nonsignificant when one's current employment status is controlled (results not shown here).

Table 7 also presents results from three other models, models 6, 7, and 8, which include, respectively, respondent's individual income, husband's income, and both. Different from the case of occupational effects, these models show that the patterns of income effects are much more clear cut. The results from model 6 show that the effects of individual income on the contrast of latent class 2 versus 1 are very strong; they are in fact the strongest effects in terms of the size of the *Z*-value among all significant effects on this contrast in the model. The effects show that compared with women with relatively low, but nonzero, individual income of less than 3.5 million yen (about \$30,000 in 1995), women who have higher individual income are significantly less likely to be traditional gender-role supporters than prowork gender-equality supporters, and women without any individual income are significantly more likely to be traditional gender-role supporters. Hence, generally speaking, higher individual income predicts a greater tendency to support gender equality of roles among women. As for the effects of respondent's individual income on the contrast of latent class 3 versus 1, the two coefficients are not significant. However, their difference of -0.778 [$= -0.364 - 0.414$] attains significance at the 5% level; compared with women without any individual income, women having relatively high income of over 3.5 million yen are less likely to be antiwork gender-equality supporters than prowork gender-equality supporters.

As for the effects of husband's income, the effects are not linear, but curvilinear: Compared with women having middle-level husband income of 4.5–7.5 million yen (about \$37,500–\$62,500), women having lower husband's income and those having higher husband's income *both* are more likely to be traditional gender-role supporters than gender-equality supporters. Husband's income has no effects on the contrast of the antiwork versus the prowork group among gender-equality supporters. Finally, the results from model 8 indicate that the curvilinear effect of husband's income tends to increase when we simultaneously control respondent's individual income. I defer the discussion of the implication of this finding to the summary and conclusion section.

As we have seen in the results from models 1 through 8, the effects of covariates related to the social strata of the respondent and her family are relatively clear cut for the contrast of latent class 2 (traditional gender-role supporters) versus latent class 1 (prowork gender-equality supporters). This is also the case for the effects of respondent's employment status,

which I describe below. However, these models did not clarify as much about the distinction between the antiwork and the prowork groups among gender-equality supporters, who differ sharply in their evaluation of women's work lives. Lifestyle variables that the SSM survey collected and a variable about respondent's subjective assessment of promotion opportunity at the present job are selected as a result of a search in a preliminary analysis for additional covariates that can inform us about differences between these two latent classes. Table 8 shows the results of models that include these covariates. Model 9 in table 8 adds to the baseline model (model 3) respondent's employment status as a covariate, model 10 adds to model 9 a covariate about respondent's promotion opportunity at the present job, and models 11–14, respectively add to model 10 one of the four lifestyle variables. Although lifestyle variables are significantly associated with both respondent's employment status and marital status, marital status does not have unique effects on the two contrasts, and its inclusion does not change the results. Marital status is, therefore, omitted from the models.

The results from model 9 show that compared with part-time employees, full-time employees are significantly less likely to be traditional gender-role supporters or antiwork gender-equality supporters than prowork gender-equality supporters. The group of self-employed and family workers—among whom the majority are family workers who work for family business typically with self-employed husbands or self-employed fathers in farm and retail trade—and women who are not employed are both significantly more likely to be traditional gender-role supporters than prowork gender-equality supporters.

The results from model 10 show that controlling for employment status, women whose probability assessment of promotion opportunity at the present job is "none," that is, who perceive the job to be a dead-end job in this respect, are significantly more likely to be antiwork gender-equality supporters than prowork gender-equality supporters. This variable has no effect on the contrast of traditional gender-role supporters versus prowork gender-equality supporters.

The results from models 11–14 show that all lifestyle variables have significant associations with the contrast of the antiwork versus the prowork group among gender-equality supporters at least at the 10% level. The two strongest effects are found for LS2 and LS4; among gender-equality supporters, women in the antiwork group are much *less* likely than women in the prowork group to be found among those who agree with the statements "I am concerned with acquiring and/or displaying refined tastes and manners" and "I prefer to make my present life pleasant rather than live frugally to save and make my later life better." They are

TABLE 8
MULTINOMIAL LOGIT LATENT-CLASS REGRESSION—3: THE EFFECTS OF EMPLOYMENT
STATUS AND LIFE STYLES

CONTRAST OF LATENT CLASSES	MODEL 9		MODEL 10	
	2 vs 1	3 vs. 1	2 vs. 1	3 vs. 1
I. Response variables (presentation omitted)				
II. Common predictor variables:				
1. Birth cohort				
Linear247***	-.231*	.229**	-.263*
2. Father's occupation—II (vs. blue collar and DK):				
Professional/clerical	-.550**	-1.201***	-.569**	-1.236***
Administrative/managerial	-.127	-.066	-.122	-.045
Sales/farm023	-.520*	.005	-.511*
3. Education (vs. postwar junior high, senior high, and junior college):				
Prewar	1.684**	-∞	1.603**	-∞
Postwar college	-.596*	-.749†	-.593*	-.688
4. Employment status (vs. part-time employee):				
Full-time	-.902***	-.763**	-.917***	-.602*
Self/family560*	-.141	.534*	.148
None922***	.021	.911***	.055
5. Promotion opportunity (vs. some or much):				
None055	.648*
III. Predictor variables added to model 10 (estimates for parameters of the 1st to the 5th variables are omitted from presentation):				
Model 11:				
LS1: "I have hobbies/lifework outside of my work and family lives" (vs. other): ^a				
Applies	-.244†	-.396†		
Model 12:				
LS2: "I prefer to make my present life pleasant rather than live frugally to save and make my later life better." (vs. other): ^a				
Applies	-.127	-.868***		
Model 13:				
LS3: "I try to make my associations with other people wide and extensive" (vs. other): ^a				
Applies056	-.442*		
Model 14:				
LS4 "I am concerned with acquiring and/or displaying refined tastes and manners" (vs. other): ^a				
Applies	-.039	-.750***		

^a The "other" category includes both "This statement does not apply to me" and "I am not certain."

† $P < .10$.

* $P < .05$.

** $P < .01$.

*** $P < .001$.

also less likely to be found among those who try to make their associations with other people wide and extensive and those who have hobbies/lifework outside of their work and family lives (though the last finding is at the 10% level). Thus, compared with the prowork group, the antiwork gender-equality supporters lead relatively poor social, cultural, and economic lives; that is, in their lives *all* of the following lifestyle characteristics are relatively lacking: (1) acquiring/displaying refined tastes and manners in their lives, (2) having the pleasant present life rather than living frugally and saving for the future, (3) having wide and extensive associations with other people, and (4) having hobbies/lifework to enrich their lives outside of their work and family lives.

SUMMARY AND DISCUSSION

Three latent classes, which I label "traditional gender-role supporters," "prowork gender-equality supporters," and "antiwork gender-equality supporters," are found among Japanese women. A conspicuous aspect here is that among women who support gender equality of roles in the family, there is a minority group who differs strongly from the majority group regarding whether they value women's work lives as much as men's. The multinomial logit latent-class regression models employed in this article clarified differences in characteristics among the three latent classes to a great extent.

Compared with prowork gender-equality supporters, traditional gender-role supporters are more likely to be found among (1) older birth cohorts, especially those having the prewar sex-segregated education, (2) those with low individual income, especially those without any individual income, (3) those having fathers whose occupations are not professional or clerical, (4) those with either high or low husband income rather than middle-level husband income, and (5) nonemployed, self-employed, or family workers rather than employees and part-time workers rather than full-time workers among employees. In particular, finding (4) indicates that although traditional gender-role supporters form a single latent class regarding their gender-role attitudes, they tend to be polarized into economically high and low classes with respect to husband's income.

On the other hand, compared with the majority of prowork gender-equality supporters who value women's work lives, the minority of antiwork gender-equality supporters who do not value women's work lives are more likely to be found among (1) younger birth cohorts, (2) those with nonfarm blue-collar origins with the exception of those having fathers with managerial/administrative jobs, (3) those having less than four years of postwar college education, and (4) those who are not full-time employees. Although these objective differences between the antiwork

and the prowork groups are not as strong as differences between traditional gender-role supporters and prowork gender-equality supporters, there are significant differences in subjective characterizations of (1) promotion opportunity at the present job and (2) lifestyles: compared with the prowork group, the antiwork group is characterized by a larger proportion of women who believe they have no promotion opportunity at their present jobs and a larger proportion of women whose lifestyle characteristics indicate relatively poor social, cultural, and economic lives. Hence, even though the antiwork group's own income and occupational status and husband's income do not differ significantly from those of the prowork group, their lower family background, lower education, and poorer lifestyle characteristics, as well as their tendency to devalue women's work lives, indicate that the prowork group and the antiwork group are two distinct *status groups*, according to Weber's terminology (Weber [1922] 1946), which not only have different lifestyles and values but are also hierarchically ordered, with the prowork group having the higher status. Hence, while traditional gender role supporters have within-group polarization in their class situation, gender-equality supporters include two distinct latent groups that differ in their status-group situation among Japanese women.

Why did the group of antiwork gender-equality supporters increase in postwar Japan? Although this is a speculation, the origin of this latent class is likely the slow historical postwar improvement in gender equality of occupational opportunity despite the rapid historical expansion of women's employment as employees (rather than self-employed or family workers), coupled with an increase in the number of holders of gender-egalitarian values among women through postwar gender-egalitarian public education. This likely generated more women historically, especially among those who have social and educational disadvantages, who work without obtaining intrinsic rewards from their work experiences and who consequently come to hold an attitude that devalues the importance of women's work lives compared with men's, despite their otherwise supportive attitudes toward gender equality of roles.

Although this article's substantive focus is not comparative in nature, it seems worthwhile to make a few comparisons between Japan and the United States. The most significant commonality is the general historical trend toward egalitarian gender-role attitudes, which was consistently reported for the American men and women during the 1960s and 1970s (Cherlin and Walters 1981; Mason, Czajka, and Arber 1976; Thornton and Freedman 1979)—though the items surveyed were different and, therefore, the content of these studies is not exactly comparable to Japanese results. An area of commonality as well as difference involves the

effects of husband-wife income difference. We have seen in the results from table 7 that when respondent's individual income is controlled, the effects of both high and low husband's income in increasing the proportion of traditional gender role supporters becomes stronger. Given that husband's income is, on average, much higher than wife's income in Japan, a high husband's income for a fixed wife's income indicates a greater husband-wife income difference and low husband's income indicates a smaller husband-wife income difference. Hence, the results show that (1) when husband's income is relatively high, a greater husband-wife income difference predicts a greater proportion of traditional gender-role supporters, but (2) when husband's income is relatively low, a smaller husband-wife income difference predicts a greater proportion of traditional gender-role supporters. In the American studies, a greater income difference between husband and wife, or a greater relative contribution by the husband to household income, leads to a greater likelihood of supporting traditional gender roles in both attitude and behavior such as the household division of labor (e.g., Erickson, Yancey, and Ericksen 1979; Hardesty and Bokemeier 1989; Plutzer 1988). No American study has reported the curvilinear effect of husband's income on gender-role attitudes. This study, however, suggests that in Japan married women with low-income husbands tend to regard the traditional gender division of labor, where the husband is the breadwinner of the family, as the state that ought to exist, as the extent of deviation from this "ideal" division of labor increases with lower husband's income and consequently greater dependence of the family on wife's income.

Finally, I would like to remark on the benefits of using multinomial logit latent-class regression analysis in general. As illustrated in this article, such analysis can identify correlates of a meaningful latent division of the population with respect to responses to a set of related observed variables. The regression framework of the analysis permits the estimation of the unique effects of predictors controlling for other variables and also enables the use of hierarchical models to see, at least descriptively, whether the effects of certain predictors change with the inclusion/exclusion of other predictors.

The latent-class characterization of response variables was particularly useful in the present application because the characteristics of the three latent classes, where attitude regarding the evaluation of women's work lives becomes a dividing factor only among gender-equality supporters, are difficult to capture with a factor-analytic approach concerned with the clustering of variables rather than the distinction of response patterns, as I stated in the introduction. However, the contrasts among categories of the latent-class variable could be formulated using continuation logit

such that covariates predict $\log(P_2/[P_1 + P_3])$ and $\log(P_3/P_1)$ instead of $\log(P_2/P_3)$ and $\log(P_3/P_1)$. The reason is that in the continuation-logit formulation the first contrast identifies the predictors of the traditional-versus-nontraditional dimension, and the second contrast identifies the predictors of the antiwork-versus-prowork dimension within nontraditional attitudes, which may seem more desirable for some researchers. The major merit of the employed multinomial link function compared with the continuation logit link function is that only the former is a straightforward extension of log-linear latent-class models. However, models with the latter link function, as well as those with the former, can be applied using the DNEWTON program (though the model specification becomes more complicated for the latter) by extending the idea described by Mare (1994), for the use of this program for continuation logit contrasts among observed categories, to contrasts among latent categories (see app. A).

As I stated in the introduction, the basic reason for using the latent variable, whether it is discrete or continuous in nature, as the dependent variable is our belief that the theoretical concept of interest cannot be measured directly by observable variables, and its measurement becomes more reliable, and theoretically more adequate, with the use of multiple indicators assuming measurement errors. This applies to many attitudinal variables that sociologists study. In addition, the use of a discrete latent variable, that is, the latent-class variable, opens the possibility of applying the latent-variable model to different kinds of dependent variables. Latent-class analyses have been used to identify distinct patterns of behavioral characteristics or life experiences, such as social-mobility patterns (Clogg 1981; Formann 1992), job characteristics (Berkelund et al. 1996), and patterns of intergenerational exchange (Hogan et al. 1993). Various possibilities thus exist for applying the latent-class model to other variables such as those related to patterns of criminal careers (D'Unger et al. 1996), patterns of underemployment, and patterns of drug-use progression (Collins and Wugalter 1992). The analyses of those variables can all be based on latent-class regression models that employ many predictors of theoretical interest to test substantive hypotheses. The method described in this article can also be extended to a longitudinal framework that employs latent-class variables (Hagenaars 1990) and to causal models where the latent-class variable with a set of predictors of its own also becomes an intervening variable that predicts another variable (Yamaguchi and Wang 1999). Hence, sociological research will greatly benefit from various substantive applications of the multinomial logit latent-class regression models described in this article and their extensions.

APPENDIX A

The Use of *DNEWTON* Program for Multinomial Logit Latent-Class

REGRESSION MODELS

The following is a description of the use of the *DNEWTON* program with multinomial logit latent-class models. A description of more general use is given by the text file *READ.LAT* that comes with the program.

The *DNEWTON* program assumes an underlying table $n(i, j)$, where row i here represents either (1) each individual when individual data are used or (2) each distinct combination of covariate states *with a nonzero frequency of observation* when cross-classified frequency data are used. In case 2, for example, model 2 of table 6 has 192 [= $4 \times 8 \times 6$] distinct combinations of three categorical covariates (birth cohort, father's occupation, and own education), but only 121 combinations have nonzero frequencies of observation. Hence, the number of rows is 121. Note that whether the frequency is zero or nonzero here implies the frequency that does *not* involve any cross-classification with the response variables. If we use individual data, the number of rows is equal to the number of sample subjects. The model assumes that each row is an independent multinomial sample.

Column j of table $n(i, j)$ indicates in the present application the distinct combination of the states of the response variables and the states of the latent-class variable. For example, for all models of tables 6–8, the number of columns is 96 [= $2^5 \times 3$] because the five response variables have two categories and the latent-class variable has three categories.

For model is given as follows:

$$\log(F_{ij}) = a_i + \sum_{k=1}^K b_k X_{ijk},$$

where F_{ij} is the expected frequency for row i and column j (or P_{ij} the expected probability in case of the individual data), a_i is the row-specific intercept, which is assumed to be present but is not estimated, X_{ijk} is the value of the k th variable for the cell (i, j) , and K is the number of parameters. Here variables X include both (1) those related to the main effects of the response variables and the latent-class variable and the interaction effects between them and (2) those related to the effects of covariates on response variables when such direct effects are hypothesized in the model). For each model in tables 6–8, the number of variables in the first category is 17, including 5 for the main effects of the response variables, 2 for the main effects of the latent-class variable, and 10 for the interaction effects

between response variables and the latent-class variable. The number of variables in the second category depends on the model. In the case of model 2 of table 6, the number is 26 [$2(1 + 5 + 7)$] for the effects of birth cohort, father's occupation, and own education on two contrasts. Hence, the total number of parameters K is 43 [$= 17 + 26$] for model 2 of table 6.

The DNEWTON program requires the following input data in free format in the following specified order: (1) the title of the model, (2) nine specific parameters, such as the total number of cells of $n(i, j)$ and the number of parameters to be estimated (see the text file READ.LAT attached to the program for details), (3) the values of the covariates X_{ijk} , (4) variable labels, (5) the array position of $n(i, 1)$, (6) data on observed frequencies and their correspondences to the cells of table $n(i, j)$, and (7) initial estimates of parameters.

The sixth group of input data is worth describing further. The DNEWTON program is generally applicable to *incompletely cross-classified data* where one-to-many correspondences exist between observed frequencies and the cells of table $n(i, j)$. The observed frequencies here are made of cross-classifications of the row and the categories of the response variables. In a latent-class analysis with L classes, each observation corresponds to L cells of table $n(i, j)$ because we assume that there are L latent frequencies, which divide each observed frequency into its latent-class components. The DNEWTON input data for each observed frequency consists of a pair of lines. The first line of the pair receives three numbers, F_{im} , i , and $n(c)$, where F_{im} is the m th observed frequency of row i , and $n(c)$ is the number of corresponding columns. When the multinomial logit latent-class regression model assumes L latent classes, each observed frequency corresponds to L columns and, therefore, $n(c) = L$. For all models of tables 6–8, $n(c) = 3$ and the last m is 32 because each observation corresponds to a distinct combined state of the five response variables. If we use individual data, F_{im} is zero except for one m where $F_{im} = 1$ for each person i —unless a sampling weight is applied. The second line of the pair receives the column numbers that correspond to the F_{im} . If we have three latent classes and use the latent-class variable as the *last* variable to be cross-classified with the response variables for ordering columns—where the columns are made of cross-classified states of the response variables and the latent-class variable—we always have three consecutive numbers here, such as “1, 2, 3,” “4, 5, 6,” and “ $J-2, J-1, J$,” depending on the corresponding observation. In the case of model 2, the first 32 pairs have entries “ $F_{1m}, 1, 3$ ” in the first line and entries in the second line changing from “1, 2, 3” to “94, 95, 96”; the second 32 pairs have entries “ $F_{2m}, 2, 3$ ” in the first line and entries in the second line changing from “1, 2, 3” to “94, 95, 96”; entries continue in this manner through the 121st 32 pairs of lines.

The DNEWTON program that one obtains may have small preassigned dimensions for the two arrays WORK and IP, such as "work (80000)" and "ip (3000)," and corresponding constants to check their maximum-allowed dimensions "nw = 80000" and "nw1 = 3000"; if this is the case, the program will give an error message if one applies any model in this article because the models use sets of input data that are too huge to be stored in these arrays. In order to eliminate this problem, simply modify the source program to expand the dimensions of the two arrays, such as "work (1600000)" and "ip (60000)," reset the nw and nw1 values such that "nw = 1600000" and "nw1 = 60000," and recompile the program to make it executable by using a FORTRAN compiler that enables the use of the extended RAM over 640K. As long as the newly assigned array dimensions do not exceed the maximum allowed by the computer's RAM, one can apply models of any size.

The use of the continuation logit link function, rather than the multinomial link function, for the latent-class variable can also be modeled by the program DNEWTON by extending the idea described by Mare (1994). For example, when there are three latent classes whose probabilities are P_1 , P_2 , and P_3 and we wish to model $\log(P_1/[P_2 + P_3])$ and $\log(P_2/P_3)$, we assume four rather than three latent classes by assuming a subdivision of P_1 , P_{12} , and P_{13} , such that $P_1 = P_{12} + P_{13}$, and imposing a constraint $P_{12}/P_{13} = P_2/P_3$, or, equivalently, independence between the row and the column of the 2-by-2 table $\{q_{ij}\}$ made of $q_{11} = P_{12}$, $q_{12} = P_{13}$, $q_{21} = P_2$, $q_{22} = P_3$. Then $\log(P_1/[P_2 + P_3]) = \log(P_{12}/P_2) = \log(P_{13}/P_3)$ holds and, therefore, we can formulate this continuation logit contrast as if it were a special multinomial logit contrast.

APPENDIX B

Some Additional Latent-Class Analyses: Tests of Equality of Response

PROBABILITIES

For the three-class model, we conducted tests for equality of response probabilities between latent classes. Equality constraints are meaningful for the conditional probabilities of a response variable between latent classes whose response probabilities for the variable are similar. From the results of table 4, we can see that models with an equality constraint for Q4 or Q5 are meaningful between latent classes 1 and 2, and those with an equality constraint for Q1, Q2, Q3, or Q4 are meaningful between latent classes 1 and 3. The results of these models are summarized in table A1. If we test equality individually, differences are nonsignificant only for the response probabilities of Q2 and Q3 between latent classes L1 and L3, and other differences are all significant. The model that combines

TABLE A1
RESULTS OF LATENT-CLASS MODELS WITH EQUALITY
CONSTRAINTS

	G ²	df	P
A. Test of equality of response probabilities between L1 and L2:			
Q4	5.20	1	<.05
Q5	7.39	1	<.01
B. Test of equality of response probabilities between L1 and L3:			
Q1	4.06	1	<.05
Q238	1	>.50
Q3	1.73	1	>.10
Q4	8.45	1	<.01
Q2 and Q3	1.97	2	>.30

equality constraints for Q2 and Q3 between L1 and L3 is most parsimonious. We did not, however, impose any equality constraints in the regression analysis because the use of covariates usually makes parameter estimates more stable and more likely to be significant—though the results remained basically the same qualitatively in the present analysis, as shown in the results of table 6.

REFERENCES

- Bandeem-Roche, Karen, Diana L. Miglioretti, Scott L. Zeger, and Paul J. Rathous. 1997. "Latent-Variable Regression for Multiple Discrete Outcomes." *Journal of the American Statistical Association* 92:1375–86.
- Berkelund, Gunn E., Leo A. Goodman, and David Rose. 1966. "The Latent Structure of Job Characteristics of Men and Women." *American Journal of Sociology* 102: 80–113.
- Brinton, Mary. 1990. "Intrafamilial Markets for Education in Japan." pp. 307–30 in *Social Institutions*, edited by M. Hechter, K. Opp, and R. Wippler. New York: Aldine DeGruyter.
- Cherlin, Andrew, and Pamela B. Walters. 1981. "Trend in United States Men's and Women's Sex-Role Attitudes: 1972–1978." *American Sociological Review* 46:453–60.
- Clogg, Clifford C. 1981. "Latent Structure Models of Mobility." *American Journal of Sociology* 86:836–68.
- . 1988. "Latent-Class Model for Measuring." Pp. 173–205 in *Latent Trait and Latent Class Models*, edited by R. Langeheine and J. Rost. New York: Penum.
- Clogg, Clifford C., and Leo A. Goodman. 1984. "Latent Structure Analysis of a Set of Multidimensional Contingency Tables." *Journal of the American Statistical Association* 79:762–71.

- . 1985. "Simultaneous Latent Structure Analysis in Several Groups." Pp. 81–110 in *Sociological Methodology*, edited by N. B. Tuma. San Francisco: Jossey-Bass.
- Collins, L. M., and S. E. Wugalter. 1992. "Latent Class Models for Stage-Sequential Dynamic Latent Variables." *Multivariate Behavioral Research* 27:131–57.
- Dayton, C. Mitchell, and George B. Macready. 1988. "Concomitant-Variable Latent-Class Models." *Journal of the American Statistical Association* 83:173–78.
- D'Unger Amy, Kenneth C. Land, Patricia L. McCall, and Daniel S. Nagin. 1998. "How Many Latent Classes of Delinquent/Criminal Careers? Results from Mixed Poisson Regression Analyses." *American Journal of Sociology* 103:1593–630.
- Ericksen, Julia A., William L. Yancey, and Eugene P. Ericksen. 1979. "The Division of Family Roles." *Journal of Marriage and the Family* 41:301–13.
- Formann, Akton K. 1992. "Linear Logistic Latent-Class Analysis for Polytomous Data." *Journal of the American Statistical Association* 87:476–86.
- Goodman, Leo A. 1974a. "The Analysis of Systems of Qualitative Variables when Some of the Variables Are Unobservable. Part I: A Modified Latent-Structure Approach." *American Journal of Sociology* 79:1179–259.
- . 1974b. "Exploratory Latent-Structure Analysis Using Both Identifiable and Unidentifiable Models." *Biometrika* 61:215–31.
- Haberman, Shelby. 1979. *Analysis of Qualitative Data*, vol. 2. New York: Academic Press.
- . 1988. "A Stabilized Newton-Raphson Algorithm for Log-Linear Models for Frequency Tables Derived by Indirect Observation." Pp. 193–221 in *Sociological Methodology* 1988, vol. 18. edited by C. C. Clogg. Washington D.C.: American Sociological Association.
- Hagenaars, Jacque A. 1990. *Categorical Longitudinal Data*. Newbury Park, Calif.: Sage.
- Hara, Junsuke, and Keiko Hiwano. 1990. "Seiteki yakuwari ishiki to shufu no chii hyouka (Gender-role consciousness and the status of housewife)." Pp. 165–86 in *Gendai Nihon no shakai kousou* (Social stratification in contemporary Japan), vol. 4. Edited by H. Okamoto and M. Naoi. Tokyo: University of Tokyo Press.
- Hardesty, Constance, and Janet Bokemeier. 1989. "Finding Timing and Making Do: Distribution of Household Labor in Nonmetropolitan Marriages." *Journal of Marriage and the Family* 51:253–63.
- Heinen, Ton. 1996. *Latent Class and Discrete Latent Trait Models*. Thousand Oaks, Calif.: Sage.
- Hogan, Dennis, David, J. Eggbeen, and Clifford C. Clogg. 1993. "The Structure of Intergenerational Exchanges in American Families." *American Journal of Sociology* 98:1428–58.
- Jöreskog, Karl G., and Arthur S. Goldberger. 1975. "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable." *Journal of the American Statistical Association* 70:631–69.
- Mare, Robert D. 1994. "Discrete-Time Bivariate Hazards with Unobserved Heterogeneity: A Partially Observed Contingency Table Approach." Pp. 341–83 in *Sociological Methodology* 1994, vol. 24. Edited by Peter V. Marsden. Oxford: Blackwell.
- Mare, Robert D., and Christopher Winship. 1991. "Loglinear Models for Reciprocal and Other Simultaneous Effects." Pp. 236–62 in *Sociological Methodology* 1991, vol. 21. Edited by P. V. Marsden. Oxford: Blackwell.
- Mason, Karen O., John L. Czajka, and Sara Arber. 1976. "Change in U.S. Women's Sex-Role Attitudes, 1964–1974." *American Sociological Review* 41:573–96.
- McCutcheon, Allan L. 1987. *Latent Class Analysis*. Newbury Park, Calif.: Sage.
- . 1996. "Multiple Group Association Models with Latent Variables: An Analysis of Secular Trends in Abortion Attitudes, 1972–1988." Pp. 79–101 in *Sociological Methodology* 1996, vol. 26. Edited by A. Raftery. Oxford: Blackwell.

- Muthén, Bengt. 1983. "Latent Variable Structural Equation Modeling with Categorical Data." *Journal of Econometrics* 22:73–78.
- Plutzer, Eric. 1988. "Work Life, Family Life, and Women's Support of Feminism." *American Sociological Review* 53:640–49.
- Sorifu Fujinmondai Tantoushitsu (Women's issue department, the office of the prime minister, Japanese government). 1984. *Fujin no seikatsu to ishiki—kokusai hikaku chousa kekka hokoku* (Women's lives and consciousness—report of the results of the international comparative survey). Tokyo: Government Printing Office.
- Thornton, Arland, and Deborah Freedman. 1979. "Change in the Sex Role Attitudes of Women, 1962–1977: Evidence from a Panel Study." *American Sociological Review* 44:831–42.
- Weber, Max. [1922] 1946. "Class, Status, Party." In *From Max Weber: Essays in Sociology*, edited and translated by H. H. Gerth and C. Wright Mills. New York: Oxford University Press.
- Winship, Christopher, and Robert D. Mare. 1989. "Loglinear Models for Missing Data: A Latent-Class Approach." Pp. 331–68 in *Sociological Methodology 1989*, vol. 19. Edited by C. C. Clogg. Oxford: Blackwell.
- Yamaguchi, Kazuo, and Denise B. Kandel. 1997. "The Influences of Spouses' Behavior and Marital Dissolution on Marijuana Use: Causation or Selection?" *Journal of Marriage and the Family* 59:22–36.
- Yamaguchi, Kazuo, and Yantao Wang. 1999. "Gender-Role Attitude, Wives' Employment, and the Household Division of Labor among American Families." Working Paper Series 99-10. University of Chicago, Alfred P. Sloan Working Family Center.

Commentary and Debate

To conserve space for the publication of original contributions to scholarship, the comments in this section must be limited to brief critiques. They are expected to address specific errors or flaws in articles and reviews published in the *AJS*. Comments on articles are not to exceed 1,500 words, those on reviews 750 words. Longer or less narrowly focused critiques should be submitted as articles. Authors of articles and reviews are invited to reply to comments, keeping their replies to the length of the specific comment. The *AJS* does not publish commenters' rebuttals to authors' replies. We reserve the right to reject inappropriate or excessively minor comments.

THE DEVALUATION OF WOMEN'S WORK: A COMMENT ON TAM¹

Why do predominantly female occupations pay less than jobs containing more men? In a recent article, Tam ("Sex Segregations and Occupational Gender Inequality in the United States," *American Journal of Sociology* 102 [1997]: 1652–92) juxtaposes two major explanations of this common finding. The "devaluation thesis" refers to a "general cultural devaluation of women's labor" (p. 1654) that leads to low wages for male and female workers in occupations filled largely by women. Tam questions the evidence for the devaluation thesis. He favors the "specialized human capital hypothesis." In this view, because of self-selection, discrimination, or both, women are concentrated in jobs that require less training that is occupation-, industry-, or firm-specific. Because specific capital is less versatile, one loses its benefits with shifts of occupation, industry, or firm. Thus, he reasons that employers will have to provide higher returns for specific than for general human capital to compensate workers for taking a greater

¹ We thank Mike Hout, Dave Jacobs, Lowell Hargens, and Reeve Vanneman for useful comments. We thank Tony Tam for making the data set used in his analysis available to us. Address correspondence to Paula England, Department of Sociology, University of Pennsylvania, 3718 Locust Walk, Philadelphia, Pennsylvania 19104-6298. E-mail: PEngland@pop.upenn.edu

Permission to reprint a comment printed in this section may be obtained only from the author.

risk. In short, in Tam's view, women's jobs pay less because they require less specific human capital, not because of a devaluation of women's work.

To test between these explanations of the low pay of "female" occupations, Tam uses cross-sectional data from the 1988 Current Population Survey (CPS). In assessing the effect of the %female of an occupation on wages, he controls for standard vocational preparation (SVP). This is a variable from the *Dictionary of Occupational Titles (DOT)* measuring how many years it generally takes to get the specific training needed for this occupation, whether the training is acquired in school, at work, or in vocational training. This is his measure of specific capital. He also controls for industry, via dummy variables. With these controls, he finds no statistically significant effect of the %female in the worker's occupation for white women, white men, black men, or black women. Thus, he argues that past research claiming a nontrivial negative effect of being in a more heavily female occupation was based on misspecified models. He criticizes the use by past researchers of extensive control variables that contain measurement error and are correlated with each other. He also criticizes the use of fixed-effect models in past research. While we are more sympathetic to fixed-effects models than Tam, we will not pursue that point here.

Using the same data used by Tam, we show that the addition of just one crucial control variable measuring occupations' demand for general education completely changes the results and restores the conclusion that there is a wage penalty for working in occupations with a higher %female. Thus, while the specific capital thesis may have some explanatory power, we believe that Tam was in error to reject the devaluation thesis.

Like Tam, many authors have used measures from the *DOT* as controls when assessing the effect of occupations' sex composition on their wages. Virtually every study using *DOT* variables has included general educational development (GED), which measures the typical requirement of the occupation for schooling that is *not* vocationally specific (as is, e.g., an engineering degree or a typing class), but is general in its relevance to many jobs (e.g., the literacy gained in high school or reading comprehension gained from a college degree). In contrast, SVP, which Tam uses, measures vocationally specific preparation, whether obtained in school or on the job. The GED and SVP are highly correlated (.74 and .82, respectively, for white women and men in these data), and they both load heavily on the first factor to come out of factor analyses using many *DOT* variables (Cain and Treiman 1981; Parcel and Mueller 1989; England 1992, chap. 3; Kilbourne et al. 1994).

However, the two measures, SVP and GED, correlate quite differently with sex composition. In these data, for white women, the correlation between GED and %female in one's occupation is -0.12 , whereas SVP and %female are correlated -0.38 . For men, the analogous correlations are

+0.17 for GED and -0.073 for SVP (results not shown). Thus, despite the high correlation between the two variables, they are clearly not redundant, because heavily female occupations are much more disadvantaged on SVP than on GED. Nor is GED, a measure of the occupation's requirement for general education, redundant with the actual education of incumbents in the occupation, which is controlled in Tam's and our models. In these data, the correlation between an individual's years of education and GED is less than 0.55 for both women and men (results not shown). Therefore, it makes sense to control for both GED and SVP in models designed to assess the net effect of sex composition on earnings.

Tam includes SVP in his models because it comes closer than any other available measure to operationalizing the concept of specific capital. We agree with this inclusion. He argues against including GED along with SVP for two reasons. First, he argues that where there is measurement error, as is surely the case for SVP and GED, multicollinearity may aggravate any bias in coefficients resulting from measurement error. While he sees this as particularly problematic for fixed-effects models, he states that it "can be devastating even in a cross-sectional model" (p. 1662) and thus omits GED and many other standard control variables from the cross-sectional models he presents.

We believe that omitting GED is unwise and has the effect of biasing the coefficient on %female, which most studies estimate to be negative, upwards toward 0. We have two reasons for this, one theoretical and one statistical. First, human capital theory, on which Tam draws, predicts returns to *both* general and specific training. Even if he is correct in his theoretical reasoning that specific capital will have a higher return, neither Tam nor any other theorist of human capital that we know of has suggested that there is *no* return to general human capital. Thus, there is no theoretical warrant for excluding GED, the best available measure for the extent to which an occupation requires types of schooling that have relevance across many firms, occupations, or industries.

Our second argument is statistical. Tam (citing Griliches 1986) points out that when one variable is measured with error, bias in coefficients caused by this may be transmitted to coefficients of correlated variables as well. We agree with Tam that *DOT* variables contain substantial measurement error. However, his cure may be worse than the disease. He simply omits GED, presumably because of its high correlation with SVP. But we do not really care about the coefficients on GED or SVP in their own right; we simply want them to function as controls to produce an unbiased coefficient on occupational %female. Since occupations containing a high proportion of women score lower (relative to occupations containing many men) on SVP than GED, given the high correlation between the two, if there really is a negative effect of occupations' %female

on their wages, the omission of GED is likely to bias the coefficient on occupational %female toward zero. We believe that this is what happened in Tam's analysis.

The general statistical point is that many experts do not think one should drop variables when confronted with multicollinearity. Indeed, most texts counsel the opposite. This is because if the two collinear control variables are both correlated with the dependent variable, then leaving one of the collinear variables out of the equation guarantees omitted variable bias (Hanushek and Jackson 1977, p. 88; Johnston 1984, p. 262; Johnston and DiNardo 1997, p. 250). If the two collinear variables are not perfectly correlated, and are correlated differently with the independent variable of interest (here occupational %female), then the coefficient on the variable of interest is sure to be biased if one of them is left out. The usual econometric view is that, since multicollinearity causes instability of parameter estimates from sample to sample, but does not bias coefficients, it is often better to do nothing than to exclude a variable when you know that doing this will cause omitted-variable bias, as is the case here. Kennedy (1985) and Studenmund (1997) counsel that leaving both variables in is particularly important when both of the collinear variables take on a significant coefficient, as we will see is the case here for SVP and GED.

The authors cited above may not have been taking into account the special problems when multicollinearity is combined with substantial measurement error. However, since we have strong reasons to suspect that omitting GED biases the coefficient on %female toward 0, whereas we do not know in what direction (or even if) the measurement error on SVP and GED cause bias in the coefficient of interest (occupational %female), we disagree with Tam's decision to leave out GED. Reasonable people can disagree on the best procedure in a world of imperfect data, but we think it important for the research community interested in the substantive question to know that simply adding one control variable—GED—to Tam's models changes the conclusion one draws from them. Our analysis below adds this control for occupations' requirement for general education to Tam's models, and the negative and significant effect of %female is generally present, supporting the devaluation thesis.

Data and Methods

We begin with a replication of Tam's analysis and then add GED to his models. We utilize the same data set (based on the May 1988 CPS) in which a measure of *occupational sex composition*, the proportion female in the occupation, from the NBER Annual Merged CPS file for 1988 is appended to individuals' record. (Tam provided his data file to us.) Spe-

TABLE 1

REGRESSION OF LOG HOURLY EARNINGS ON SEX COMPOSITION, SPECIFIC VOCATIONAL PREPARATION, GENERAL EDUCATIONAL DEVELOPMENT, AND SELECTED CONTROL VARIABLES

	WHITE WOMEN		WHITE MEN	
	Replication (1)	Extension (2)	Replication (3)	Extension (4)
SC	-.007 (.018)	-.056* (.018)	.002 (.019)	-.080* (.020)
SVP078* (.004)	.040* (.005)	.069* (.003)	.024* (.004)
Industry	Yes	Yes	Yes	Yes
GED101* (.008)		.113* (.008)
R ²496	.505	.532	.541
N	9,265	9,265	10,372	10,372

NOTE.—Other variables included in the analyses are South, West, Midwest, SMSA, veteran, union, education, experience and its square, tenure and its square, and full-time. SEs are given in parentheses.
* $P < .05$, two-tailed test.

cific vocational preparation from the *Dictionary of Occupational Titles* is also appended to each record (England and Kilbourne 1989). We extend Tam's analysis by including a measure of general educational development from the *DOT* (England and Kilbourne 1989). Our tables show Tam's results (column marked "Replication"); we were able to replicate his *N*s and coefficients precisely. Our tables also show our extension that adds GED to the models (column marked "Extension").

Findings

Each of our numbered results below correspond to Tam's numbered results.

RESULT 1.—*The negative effect of occupational %female on wages is statistically and substantively significant among white women once general training is added to Tam's model, which includes specialized training and a set of basic control variables.*²

Tam found no significant net sex composition effect among white women (table 1, replication). However, after adding GED, the sex compo-

² The basic controls referred to are those used by Tam, consisting of dummy variables for SMSA/non-SMSA, region, veteran status, union membership, and full-time work, as well as continuous measures of schooling, potential work experience, and tenure.

sition of the occupation has a significant negative effect on white women's earnings (table 1, extension). Sex composition is measured as a proportion, so the coefficient of -0.056 indicates that movement from an all male to all female occupation would lower earnings by 5.6%. While *some* (about two-thirds) of the relationship between sex composition and wages from a model with just the basic individual-level controls ($b = -0.158$, results not shown) is "explained away" by adding SVP, GED, and industry variables, the effect of sex composition remains significant.

RESULT 2.—*The negative effect of occupational %female on wages is statistically and substantively significant among white men as well.*

Tam found a nonsignificant effect of sex composition on white men's wages after controlling for industry and SVP (table 1, replication). When we add the control for GED, the estimate of the sex composition effect becomes negative (-0.080) and statistically significant. Men moving from an all-male to an all-female occupation would lose 8%, *ceteris paribus* (table 1, extension).

RESULT 3.—*The negative effect of occupational %female on wages does not differ significantly by race among either men or women. Thus, it is safest to conclude that both blacks and whites are penalized for working in more female occupations.*

Tam found no significant effect of sex composition for African-American men and women (table 2, replication). Our extension, adding the control for GED, still finds a nonsignificant effect of occupational %female for black men and black women. Since the devaluation thesis implies that workers, regardless of race or sex, are hurt by working in a predominantly female occupation, one could argue that if the effect does not hold quite generally, this is an argument against the thesis. But as we divide into smaller subgroups, *Ns* get smaller, so effects might lose significance for this reason alone. Thus, it seems to us that subgroup analyses should proceed by testing whether groups show differences in the coefficient of interest (the effect of sex composition) that are statistically significant. If not, analyses pooled by race make more sense.³

Following this logic, to test whether the effect is significantly different for blacks than whites, in results not shown, we pooled black and white women, added a dummy variable for whether an individual was black, and interacted this with all other independent variables in the analyses in tables 1 and 2. We then did the analogous race-pooled analysis for men. Whether we used our preferred specification that adds GED or Tam's specification that excludes GED, we did not find the effect of occupational

³ One could argue for pooling by sex as well, but we have retained separate analyses by sex.

TABLE 2

REGRESSION OF LOG HOURLY EARNINGS ON SEX COMPOSITION, SPECIFIC VOCATIONAL PREPARATION, GENERAL EDUCATIONAL DEVELOPMENT, AND SELECTED CONTROL VARIABLES

	BLACK WOMEN		BLACK MEN	
	Replication (1)	Extension (2)	Replication (3)	Extension (4)
SC051 (.056)	-.020 (.058)	.014 (.059)	-.019 (.061)
SVP094* (.012)	.045* (.016)	.079* (.010)	.058* (.014)
Industry	Yes	Yes	Yes	Yes
GED111* (.025)		.054* (.028)
R^2498 ^a	.507	.516	.518
N	1,042	1,042	855	855

NOTE.—Other variables included in the analyses are South, West, Midwest, SMSA, veteran, union, education, experience and its square, tenure and its square, and full-time. SEs are given in parentheses.

^a We were able to exactly replicate the coefficients and N for this analysis; however, our adjusted R^2 does not match that reported in Tam's published table (.491).

* $P < .05$, two-tailed test.

%female to be significantly different for black than for white women, or to be significantly different for black than for white men. We then ran models for each sex, still pooling the two racial groups and retaining the significant but omitting the nonsignificant interactions with race. These models showed significant negative effects of occupational %female for each sex when GED was present in the model (although not when GED was absent; results not shown). Thus, since occupational %female has a negative effect for each sex in pooled-race analyses, an effect that is not significantly different for blacks than for whites, we think the most sensible conclusion (for those who agree that it is appropriate to control for GED) is that there is evidence of devaluation for both blacks and whites.

RESULT 4.—*There is one subgroup among whom the devaluation thesis does not hold—white women working part-time; even with GED controlled, occupational sex composition has no significant effect on their pay.*

Tam examined the effects of sex composition separately for full-time and part-time white and African-American women to see if evidence of devaluation was robust across subgroups. Tam's analysis showed nonsignificant positive effects of %female for both full- and part-time black women, significant positive effects for part-time white women, and a sig-

TABLE 3

REGRESSION OF LOG HOURLY EARNINGS ON SEX COMPOSITION, SPECIFIC VOCATIONAL PREPARATION, GENERAL EDUCATIONAL DEVELOPMENT, AND SELECTED CONTROL VARIABLES

	REPLICATION		EXTENSION	
	Full-time	Part-time	Full-time	Part-time
White Women:				
SC	-.044*	.084*	-.082*	.017
	(.020)	(.040)	(.020)	(.043)
SVP068*	.109*	.031*	.075*
	(.004)	(.008)	(.005)	(.011)
Industry	Yes	Yes	Yes	Yes
GED104*	.079*
			(.009)	(.018)
R^2453	.440	.463	.444
N	6,793	2,472	6,793	2,472
Black Women:				
SC055	.137	-.004	.004
	(.058)	(.175)	(.059)	(.196)
SVP087*	.136*	.043*	.076
	(.012)	(.040)	(.016)	(.056)
Industry	Yes	Yes	Yes	Yes
GED105*	.112
			(.026)	(.075)
R^2486	.328	.496	.333
N	852	190	852	190

NOTE.—Other variables included in the analyses are South, West, Midwest, SMSA, veteran, union, education, experience and its square, and tenure and its square. SEs are given in parentheses.

* $P < .05$, two-tailed test.

nificant negative effect for full-time white women (table 3, replication).⁴ Thus, in his analysis, the devaluation thesis was only supported with the right sign and significance for one out of the four groups of women, and the significant *positive* effect of %female for part-time white women seemed particularly damaging to the devaluation thesis.

Our extension, which adds GED to each of the separate subgroup models, does not find a significant positive effect of occupational %female for part-time white women, or any group of women (table 3, extension). Like

⁴ Tam's text (p. 1676) refers to this effect for full-time white women as "insignificantly negative," but his table reports it to be statistically significant at the 0.05 level. He has confirmed in a private communication that statement of insignificance was a clerical error and the table is correct, consistent with our replication.

Tam, we find a significant negative effect for full-time white women. We find nonsignificant effects for part-time women, black or white, and for full-time black women. As argued previously, we did not want to conclude that effects differ across subgroups without testing for the statistical significance of such differences across subgroups. To test for group differences, in results not shown, we pooled all women into one sample. We made white full-time women the reference category and included dummies for white part-time, black part-time, and black full-time women, as well as interactions of each of these dummies with all independent variables. The interactions of occupational sex composition with the dummies were all nonsignificant except that involving white part-time women. Thus, it is only for white women working part-time that we can clearly reject the presence of a wage penalty for working in more heavily female occupations.⁵ This relatively small subgroup is somehow distinct, and future research might explore reasons for this exception to the devaluation mechanism. However, adding GED to the model changes the finding from one in which white women working part-time actually benefit from being in an occupation with more females (table 3, replication of Tam) to one in which there is no significant effect of occupational sex composition for this subgroup (table 3, extension).

Conclusion

We believe that a measure of occupations' requirements for general human capital should be added to Tam's models. When we add such a measure, GED, results generally show a wage penalty for those working in a

⁵ The alert reader might wonder, as we did, if the important variable negating the effect of occupational sex composition was part-time status, and perhaps black part-time women did not show up as significantly different from white full-time women because of their small sample size. If this were true, we might find no effect of occupational sex composition on the wages of part-time men either. The idea seemed intuitively plausible to us that, at least within skill levels, part-time work might be paid badly—but quite uniformly badly across more heavily male and female jobs. But supplementary analyses did not support this conjecture. We performed analyses for both men and women that tested whether part-time status interacted significantly with occupational %female. (These were pooled by race since the *N*s for part-time black workers are so small.) The interaction of part-time status with occupational %female was not significant for either sex (results not shown). Thus, in general, the penalty for being in a more heavily female occupation is no different among part-time workers than among full-time workers. However, our results discussed in the text show white part-time women workers to be significantly different from white full-time women workers. So, although we do not know the reason, it appears there is something unique about the experience of white part-time women workers in their immunity from the general penalty for "female work" (or, to put the same thing another way, in their failure to experience a premium for working in "male" occupations).

more female occupation.⁶ After controlling for individuals' human capital, the specific and general human capital demands of the occupation, and industry, working in an all-female occupation results in earnings that are 5%–8% less than earnings in all-male occupations. These effects are not significantly different for blacks than whites. Further division into small subgroups reveals one group with a statistically significant difference in effect: white women who work part-time have wages unaffected by occupational sex composition. Outside of this segment of the labor force, the mechanism of devaluation holds quite generally. Workers pay a penalty for working in an occupation containing more women. Or, to put the same thing another way, workers enjoy a premium for working in an occupation containing more men.

It is beyond the scope of our debate with Tam, and beyond what our data can illuminate, to specify the exact mechanisms involved in producing the effect of occupational sex composition on wages. But we offer some description of how we think this process works. At certain points in the life of organizations, particularly when they are first formed, or undergo significant restructuring, the relative wages of various jobs are somewhat up for grabs, although not unaffected by market forces. At these points, gender biases infect the decisions that managers or consultants make about jobs' wages. In part, this results from a cognitive distortion that involves underestimating the relative contribution to profits or other organizational goals of the work done in "female" jobs. In part, the bias results from generalizing to women's *jobs* the relatively low value accorded *women* as persons by our culture. Later, bureaucratic inertia takes over, setting the wage consequences of these biases in stone; the relative wage levels of jobs within an organization often remain quite constant for decades. If any attempt is made to change relative wage levels in a way favorable to female occupations, individual or collective action by male workers may bring male jobs wages above that of female jobs again. Market processes then reproduce all these effects. That is, if most employers have at some point inscribed these biases into their wage structures, this affects the "going wage" in predominantly female and male occupations that other new organizations must pay to hire workers of typical quality into the occupation. These market forces dictate that even an organization not beset by these cognitive and cultural biases will be forced to pay a higher wage to recruit in male occupations than comparable female occu-

⁶ Another approach to the problem of collinearity of GED and SVP is to make a scale combining them. In results not shown, we used the scale that emerged from a principal components factor analysis (without rotation) of the two variables in models analogous to those in tables 1 and 2. This analysis yielded effects of %female similar to those in tables 1 and 2 (results not shown).

pations. While they could choose to pay more than they need to—more than the market or average wage—in female occupations, few profit-minded employers will do so. In these ways, gender bias is perpetuated over time. Absent significant collective action or legislation to achieve comparable worth in pay scales, the social and inertial forces described above may keep the pay penalty for working in “female” jobs in force indefinitely.

REFERENCES

- Cain, Pamela S., and Donald J. Treiman. 1981. “The *Dictionary of Occupational Titles* as a Source of Occupational Data.” *American Sociological Review* 46:253–78.
- England, Paula. 1992. *Comparable Worth: Theories and Evidence*. New York: Aldine De Gruyter.
- England, Paula, and Barbara Stanek Kilbourne. 1989. *Occupational Measures from the Dictionary of Occupational Titles for 1980 Census Detailed Occupations*. ICPSR #8942 (MRDF). Ann Arbor, Mich.: Inter-university Consortium for Political and Social Research.
- Griliches, Zvi. 1986. “Economic Data Issues.” Pp. 1466–514 in *Handbook of Econometrics*, vol. 3. Edited by Zvi Griliches and Michael D. Intriligator. Amsterdam: North-Holland.
- Hanushek, Eric A., and John E. Jackson. 1977. *Statistical Methods for Social Scientists*. New York: Academic.
- Johnston, J. 1984. *Econometric Methods*. New York: McGraw-Hill.
- Johnston, J., and Dinardo, J. 1997. *Econometric Methods*. New York: McGraw-Hill.
- Kennedy, Peter. 1985. *A Guide to Econometrics*, 2d ed. Oxford: Blackwell.
- Kilbourne, Barbara Stanek, Paula England, George Farkas, Kurt Beron, and Dorothea Weir. 1994. “Returns to Skill, Compensating Differentials, and Gender Bias: Effects of Occupational Characteristics on the Wages of White Women and Men.” *American Journal of Sociology* 100:689–719.
- Parcel, Toby L., and Charles W. Mueller. 1989. “Temporal Change in Occupational Earnings Attainment, 1970–1980.” *American Sociological Review* 54:622–34.
- Studenmund, A. H. 1997. *Using Econometrics: A Practical Guide*. Reading, Mass.: Addison-Wesley.

PAULA ENGLAND

University of Pennsylvania

JOAN M. HERMSEN

University of Missouri, Columbia

DAVID A. COTTER

Union College

OCCUPATIONAL WAGE INEQUALITY AND DEVALUATION: A CAUTIONARY TALE OF MEASUREMENT ERROR¹

I welcome the commentary of England, Hermesen, and Cotter on my *AJS* article, "Sex Segregation and Occupational Gender Inequality in the United States." It gives me the opportunity to further clarify the fundamental issues raised by the quantitative testing of the devaluation hypothesis (DEVH)—the main sociological answer to the question of why occupations with a higher proportion female tend to pay less. The DEVH is a cultural theory of the wage determination mechanism that drives the correlation between wage and occupational sex composition (SC) and purportedly justifies comparable worth policies.² My article challenges the credibility of the DEVH with a methodological critique of prior research designs, a cautious reinterpretation of the old evidence, and a series of powerful new findings.

England et al. take issue with the central findings of my study. The punch line of their analysis is that the addition of just one independent variable (general educational development, GED) to my full model (1) *completely changes the results* and (2) *restores the conclusion* that there is a wage disadvantage for occupations with a higher percentage of female. They also make the stronger and more specific claims that (3) *my modeling strategy is ill-advised* and (4) my findings may well be the *results of omitted variable bias*. In other words, the omission of GED is unjustifiable, and my findings are not credible.

¹ I thank Paula England for comments on an earlier draft of this reply. This research was supported by Academia Sinica through the Organization-Centered Society Project and the Institute of European and American Studies. Direct correspondence to Tony Tam, Academia Sinica, Institute of European and American Studies, Nankang, Taipei 11529, Taiwan. E-mail: tam@sinica.edu.tw

² Gross wage effects of sex composition (percentage female) are indeed dramatic—they are often over 20%. In my Current Population Survey data, the gross SC effects among white women and black women are 25% and 35%, respectively. The gross effects among white and black men are 13% and 15%, respectively. These numbers seem to fit the notion that women and racial minorities suffer the most from devaluation of women's work. The negative impact of SC one gets from case studies is even more staggering (e.g., Acker 1989, p. 99). Not surprisingly, comparable worth policy advocates often have emphasized the severity of the gross SC effects. Their interpretation is that sex discrimination in wage determination is responsible for the gross effects. However, England et al. and others have tacitly chosen to defend a much weakened hypothesis—devaluation is responsible for a statistically significant and nontrivial portion of the gross SC effect. England et al. must have realized that much of the gross SC effect is attributed to reasonable worker differences. Only residual (i.e., conditional) SC effects may be due to cultural devaluation. *Contrary to the presumption of the policy literature, then, devaluation cannot be responsible for most of the gross SC effects.*

Before I address the four central claims of England et al., I should mention a historical fact about my *AJS* article. In the first draft submitted to the *AJS*, I had presented and extensively discussed the findings in tables 1 (the main empirical support of England et al.) and 2 of the England et al. commentary. In response to reviewer suggestions, this material was moved to a technical report (Tam 1996).³ This technical report is cited in the published paper, has been circulated among colleagues, and has been made available to all readers who have requested it. In short, England et al. and I draw different conclusions from the same set of results.

This note has three principal conclusions. First, a careful review of their tables reveals that the new support for the DEVH is weak even in the GED models, especially in light of sampling error, the remarkably coherent evidence against the DEVH in my article, and the likelihood that the GED models may be misspecified.⁴ Second, England et al. make no attempt in their reanalysis to correct for measurement error in their variables, which was one of the central issues raised in my *AJS* article. The structure of this measurement error probably induces a bias in the SC coefficient in a direction that would appear to support the DEVH. If this bias was corrected, the already weak support for the DEVH found in the naive analyses would be weaker or nonexistent. Third, the assertion that the 1997 findings arise from omitted variable bias does not follow from their regression analysis and requires additional tests of two competing interpretations: omitted variable bias and measurement error bias. I consider a stringent test of the measurement error interpretation. The results are surprisingly supportive of the measurement error interpretation of the GED results and strongly reinforce the 1997 findings. I justify these conclusions in the following sections.

³ When submitting the manuscript for a second-round review, I included the technical report for the benefit of the referees and the editorial board. The referees and the editors were satisfied with the contents of the technical report and did not suggest incorporating the report back into the paper. The bottom line is that the referees of both rounds and the former *AJS* editorial board have taken full account of my technical report and therefore the central findings of England et al.

⁴ Although England et al. argue that the inclusion of GED is theoretically justified by human capital theory, this is not so obvious to me. Human capital theory certainly postulates that general skills should earn market returns. But years of schooling rather than GED has been the most stable and widely vindicated indicator of the general human capital of a worker. Given the sloppy coding of GED, commented on in my previous article (see, p. 1685), it is difficult to imagine how it could be a good supplement to years of schooling as an indicator of a worker's general human capital. Given the way in which GED deviates from the typical years of schooling of workers in an occupation, a coder may have subjectively taken into account the typical earnings of the occupation. This could be an artificial source of strong GED effect on wage.

Have the GED Results Completely Changed the Picture?

The big picture.—There are three cornerstones of my 1997 conclusions: (1) Through a critique of prior research designs, statistical methods, and the old evidence, I specify ways in which prior research is biased in favor of the DEVH. The available evidence provides surprisingly little support for the DEVH.⁵ (2) By discovering new and striking findings that are very difficult to reconcile with the DEVH and its common defenses, I raise serious doubt about the role of devaluation. In stark contrast, *all* the old and new findings can be easily explained in terms of an alternative interpretation—the specialized human capital hypothesis (SHCH). The burden of proof should be with the DEVH. (3) In my article, the SHCH meets a much more exacting standard of empirical adequacy than that of the DEVH. The DEVH fails to be supported despite the fact that sampling error is much more likely to have produced favorable findings for the DEVH.⁶ Future research should subject both the DEVH and SHCH to stringent tests.⁷

Claims 1 and 2: complete changes of results and conclusions.—Tables 1–3 are supposedly the empirical basis of the first two claims of England et al. Tables 1 and 2 mirror three of the four central results of my 1997 study, and table 3 concerns the other central result. But what do the estimated SC effects say in these tables?

The two new estimates in table 1 (for whites) fall between -5% and -8% , consistent with the DEVH. The two new estimates in table 2 remain

⁵ Supporters of the DEVH have certainly drawn on a wealth of ethnographic and case evidence. However, the relevance of the evidence is critically undermined by a theoretical fallacy that perpetuates even among prominent economists: the prevalence of discriminatory behavior in the market implies the presence of discrimination-distorted market prices. In fact, market prices do not necessarily reflect any discriminatory action present in a market (Becker 1957). Heckman (1998, p. 102) reiterates the central economic result that prices are set at the margin: "The impact of market discrimination [on prices] is not determined by the most discriminatory participants in the market, or even by the average level of discrimination among firms, but rather by the level of discrimination at the [marginal] firms where ethnic minorities or women actually end up buying, working and borrowing."

⁶ Recall that the estimated SC effects for blacks and whites, men and women, are absolutely close to zero. *None* of the four key estimates (replications in tables 1 and 2 of England et al.) is significantly negative, even if we raise the statistical significance level to .5—that is ten times the conventional .05 level. If chance fluctuation or systematic bias is driving the estimates, we would have expected more chaos in the estimates—we would not have expected all estimates to be so consistently nonnegative, let alone so consistently close to zero. Even when the true effect is absolutely close to zero, sampling error could have produced estimates far from zero.

⁷ To be credible and relevant, supporters of the DEVH must confront rather than ignore the pitfalls of the old literature when designing empirical tests that can adjudicate between the DEVH and the SHCH.

at odds with the DEVH because they are absolutely close to zero (about -2%) and statistically insignificant. Moreover, three of the four estimates in table 3 are also absolutely close to zero (about -0.4% to 1.7%) and statistically insignificant. The big story is *five of the eight* estimated SC effects based on a model that has included GED turn out to be at odds with the DEVH. Nonetheless, England et al. confidently proclaim that these mixed results represent a complete change of my previous results.

In a world of imperfect data and models, an emphasis on consistency is a very useful safeguard against the danger of capitalizing on chance. England et al. have downplayed the extent to which the inconsistency in their reported findings undermines the DEVH. They have chosen to emphasize the three estimates seemingly favorable for the DEVH and discount the five favorable for the SHCH. Simply put, claims 1 and 2 inaccurately characterize the GED findings and ignore the danger of capitalizing on chance.⁸ Most significant, the GED analysis *bypasses* the three cornerstones that support my 1997 conclusions. The GED results provide a very weak basis for restoring confidence in the devaluation hypothesis.

Claim 3: bigger is always better.—England et al. argue that the best strategy is *always* to include more rather than fewer control variables, even when those control variables are noisy and correlated. I call this position the bigger-is-better (BIB) assumption. They use the BIB assumption to justify the assertion that their OLS regression models that include GED will produce less biased estimates of SC effects than the estimates from my models that omit GED. However, the extent to which the inclusion of a variable increases or decreases bias depends, among other things, on the covariance structure among all the variables and the structure of measurement errors.

In fact, the validity of the BIB assumption is problematic (Lieberson

⁸ Like the four central findings of my article, the relevance of results 3 and 4 of England et al. hinge on the validity of results 1 and 2. Results 1 and 2 are fundamental and thus the focus of this reply. It is worth noting, however, that England et al. have misused the interaction test results in providing support for their results 3 and 4. First, the interaction tests cannot possibly contradict the race-specific analysis in tables 2 and 3. In fact, the race-specific analysis is more informative than the interaction analysis. The race-specific analysis (especially if done by full-time/part-time status) clearly reveals that the black samples are simply too small to produce precise point estimates of the effect of percentage female. Second and most important, the large standard errors imply that the point estimates for blacks are consistent with a wide range of true values—both large negative and large positive values, not just compatible with the negative estimates for whites. The data on blacks are not rich enough for distinguishing the SC effects among blacks from those among whites. It is therefore highly misleading for England et al. to conclude that the interaction test justifies their result 3 and that it is safest to regard blacks and whites as both penalized by working in female occupations.

1985). Statisticians have investigated this problem and made explicit the conditions under which the BIB assumption will be violated with dire consequences (e.g., Griliches 1986). For our purposes, the most fundamental requirement of the BIB assumption is *the absence of measurement error in correlated independent variables*. The BIB assumption is untenable for models with noisy correlated variables. The greater the noise and the higher the correlations, the more perilous is the assumption because the more serious are the biases on standard regression estimation. As I will argue next, GED and specific vocational preparation (SVP) represent the paradigmatic case of noisy correlated variables in which measurement error cannot be ignored. The failure of England et al. to correct their GED results for measurement error undermines the validity of their conclusions. Their claim 4 is an example of the common oversight of the ruses of measurement error.

Omitted Variable Bias or Noise?

I will sketch the statistical foundation for evaluating claim 4 of England et al. and then quantitatively characterize the interplay of correlation and measurement error. The numerical results will strengthen the case against claims 3 and 4 of the England et al. analysis.

Statistical theory.—Consider a standard linear model for dependent variable Y and independent variables X_1 and X_2 : $Y = a + b_1 X_1 + b_2 X_2 + e$. The true underlying model is $Y = \alpha + \beta_1 Z_1 + \beta_2 Z_2 + \epsilon$. Assume that $X_1 = Z_1 + u$, that is, there is classical error in the indicator for Z_1 , but $X_2 = Z_2$. Further, b_1 and b_2 are the OLS estimators of β_1 and β_2 , respectively. To simplify notation and without loss of generality, scale X_1 and X_2 so that their variances equal one. For this model, asymptotic biases (the probability limit of the deviations of the estimators from the true parameters as sample size goes to infinity) are given by (Griliches 1986):

$$\text{plim}(b_1 - \beta_1) = -\beta_1 \lambda / (1 - \rho^2) \quad (1)$$

and

$$\text{plim}(b_2 - \beta_2) = \rho \beta_1 \lambda / (1 - \rho^2) = -\rho [\text{bias in } \beta_1], \quad (2)$$

where $\rho = \text{corr}(X_1, X_2)$, and $\lambda = (\sigma_u^2/\sigma_1^2)$ is a measure of unreliability (the relative amount of error variance) in the observed variable X_1 . As is well-known, the OLS estimator for the coefficient of the erroneous variable (X_1) is biased toward zero. But the simple result in equation (1) has many important implications: the bias is proportional to error variance and inversely proportional to the residual variance of X_1 (i.e., conditional on the

analysis, estimation method,⁹ and other independent variable in the model). The bias will exist even if there is *only one* other control variable in the model.¹⁰ The bias will also lead to the underestimation of the standard errors of b_1 and b_2 , and so conventional significance tests will be biased in favor of the DEVH. The result in equation (2) is also telling: the bias on the coefficient of the erroneous X_1 will be transmitted to other variables correlated with it, even if the other variables are perfectly measured. England et al. never take into account any of these consequences.

Testing Competing Interpretations.—Although the 1997 findings challenge the DEVH, England et al. do not believe in the validity of my findings. Central to our debate are two interpretations of the GED results. One interpretation is omitted variable bias. England et al. believe that the bias from omitting GED induces a positive bias on what should have been a negative b_{sc} (see, e.g., Hanushek and Jackson 1977, pp. 81–82). To them, controlling for GED should eliminate the positive omitted variable bias and restore the negative b_{sc} . The other interpretation is measurement error bias or noise. I believe that the inclusion of GED as a cure for omitted variable bias is worse than the disease. The inclusion of GED will aggravate the downward measurement error bias on b_{syp} and therefore induce a *negative bias* on b_{sc} . The GED findings are driven by the ruses of measurement error rather than the elimination of omitted variable bias.

In general, however, we do not know the size of any omitted variable bias without knowing the true parameters and the true extent of measurement errors. To empirically evaluate the two interpretations, I will reanalyze the GED model with *hypothetical* levels of noise (unreliability) for white women and men—the samples that underlie table 1 of England et al. The crucial question is *Is there any assumption of measurement error that can translate the GED results for whites into my 1997 results?*

⁹ Some common ways in which the variance of X_1 would be reduced include (1) doing subgroup analysis when a big portion of the sample variance is between groups; (2) the use of fixed-effects models for panel data, which is equivalent to throwing away interpersonal variance and focusing on intertemporal changes within individuals. Although there is nothing inherently wrong with subgroup analysis and fixed-effects models, researchers must be fully cognizant of their consequences for variables with measurement error.

¹⁰ Thus the use of factor scores to reduce the dimensionality of control variables does not really circumvent the interplay of unreliability and correlated variables. England et al.'s note 5 reports the use of factor analysis to combine SVP and GED. It could have made things worse because it forces a synthesis of the theoretically different dimensions of human capital.

If there is, the noise interpretation can perfectly explain the discrepancies between the GED results and my findings. Under the omitted variable bias interpretation, there is no reason why the estimates of different coefficients will work together this way under *any* assumption of noise. Thus if I can demonstrate that such an assumption indeed exists, the result will constitute a very strong confirmation of the noise interpretation.¹¹

Are the GED findings really compatible with $\beta_{sc} = 0$ under an assumption of measurement error? Can the GED model reproduce my 1997 estimates by taking into account measurement error? The answer is yes. In fact, the required assumption is exceedingly simple: the unreliability of SVP is modest (say, less than 0.20), but its impact dominates the effects of noise in all other independent variables (which may have minor unreliability, weak correlation with SC and SVP, and so forth). To fix ideas, the following estimation will proceed by simply ignoring any measurement error in the other independent variables. I reestimate the GED models in table 1 of England et al. and adjust for classical measurement error in SVP.¹² The results are striking: by adjusting for modest error in SVP, the b_{sc} and b_{svp} of a GED model come close to being identical to those estimated without GED (i.e., estimates in my article). For white women, adjusting for an unreliability of 0.15 in SVP will shift their estimates of b_{sc} from -0.056 to -0.007 (-0.007 was my estimate) and b_{svp} from 0.040 to 0.074 (0.078 was my estimate). Similarly for white men, adjusting for an unreliability of 0.17 is able to shift their estimates of b_{sc} from -0.080 to -0.002 (0.002 was my estimate) and b_{svp} from 0.024 to 0.071 (0.069 was my estimate).

This analysis strongly suggests that *the differences between the England et al. estimates and mine may well be due to their failure to take account of measurement error in the GED models*. The key GED parameter estimates can indeed be translated back into my 1997 estimates. After adjusting for measurement error, there is virtually nothing left in the key results of England et al. for an omitted-variable story to explain. *An error-in-SVP story is sufficient to explain the GED results*.¹³ The GED results

¹¹ It is hard to overemphasize that *this is an exceptionally stringent test for the noise interpretation*. The test by design favors findings consistent with the omitted variable bias interpretation. *Even if* the noise interpretation is valid, it should be extremely difficult for a researcher to arrive at the suitable assumption and produce findings that can exactly translate the GED results into my 1997 results.

¹² I use the program for errors-in-variable regression (EIVREG) available in Stata 6.0.

¹³ What, then, is the impact of the omission of GED on b_{sc} ? This question can be examined in two steps. The first step is to identify the direct role of SVP in producing the negative b_{sc} in the England et al. results. Equation (1) tells us that the downward

emphasized by them are evidently driven by the interplay of measurement error and correlation. The additional analysis here is inconsistent with the omitted variable bias interpretation of my 1997 results and the GED results.

Conclusion

My 1997 article may not have definitively settled the debate about the wage effects of occupational SC. However, the GED results emphasized by England et al. offer no new insight. Nor do the results restore the credibility of DEVH. The reanalysis of the GED models only strengthens my previous conclusions. The burden of proof is still with the DEVH. More sophisticated studies of the DEVH and SHCH are certainly necessary to further our understanding, but good studies should not repeat the pitfalls of the prior literature.

The analysis here also offers a general lesson. The ruses of measurement error have plagued the old literature on occupational wage inequality and may well be rampant in quantitative social science. My 1997 article and the analysis of this reply provide a cautionary tale against recklessly adding control variables to a statistical model. Producers and consumers of statistical analysis should take heed: more controls are not always preferable, bigger models may not be better.

measurement error bias on b_{SVP} can be greatly aggravated by the correlation between SVP and other control variables. Thus the addition of a strongly correlated GED could have produced a large reduction in b_{SVP} and resulted in a negative b_{SC} . To test this argument, constrain the coefficient of SVP (b_{SVP}) to be those from the England et al. results, and observe b_{SC} . (One could also do this using the coefficients of SVP from the error-in-SVP model of the preceding paragraph. The results would be practically the same.) This constraint on b_{SVP} shifts b_{SC} from a trivial -0.007 to a -0.095 (compared to -0.056 in the GED model) for white women and from a trivial 0.002 to significant -0.060 (-0.080 in the GED model) for white men. These results suggest that the negative residual SC effects among whites are driven by changes in the coefficients of SVP rather than any omitted variable bias. The second step is to assess the direct impact on b_{SC} of including GED. We can do this by comparing the results of the first step with the GED results of England et al. For white women, including GED will shift the residual SC effect from -0.095 to -0.056 . The omitted variable bias due to GED is $(-0.095) - (-0.056) = -0.039$. But this modest bias is *opposite* to the direction postulated by them. That is, contrary to the assumption of England et al., the omission of GED shifts b_{SC} toward the negative direction and, equivalently, the inclusion of GED actually makes b_{SC} more positive. For white men, the story is somewhat different. The omission of GED shifts b_{SC} from -0.080 to -0.060 , just as they hope for. However, even here, the big story is that the omitted variable bias is meager (0.02) when compared to the -0.062 bias on b_{SC} due to the correlation between GED and SVP. In any event, the evidence is contrary to the England et al. interpretation. Omitted variable bias does not consistently follow the direction postulated by them and has little impact on the b_{SC} in the GED results.

REFERENCES

- Acker, J. 1989. *Doing Comparable Worth: Gender, Class, and Pay Equity*. Philadelphia: Temple University Press.
- Becker, G. S. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press.
- Griliches, Z. 1986. "Economic Data Issues." Pp. 1465–514 in *Handbook of Econometrics*, vol. 3. Edited by Z. Griliches and M. D. Intriligator. New York: North-Holland.
- Hanushek, E. A., and J. E. Jackson. 1977. *Statistical Methods for Social Scientists*. New York: Academic Press.
- Heckman, J. 1998. "Detecting Discrimination." *Journal of Economic Perspectives* 12 (2): 101–16.
- Liebersohn, S. 1985. *Making It Count: The Improvement of Social Research and Theory*. Berkeley and Los Angeles: University of California Press.
- Tam, T. 1996. "Is Bigger Always Better? A Note on the Problem of Noisy Correlated Variables." Manuscript. Academia Sinica.

TONY TAM

Academia Sinica

Book Reviews

Department and Discipline: Chicago Sociology at One Hundred. By Andrew Abbott. Chicago: University of Chicago Press, 1999. Pp. xii+249. \$45.00 (cloth); \$18.75 (paper).

Arthur L. Stinchcombe
Northwestern University

Andrew Abbott has a quirky and brilliant mind, an oversupply of self-esteem, and patience. He can therefore explore a subject deeply, even if no one else wants to go so deep. As a result, his professions book shook up the discipline. The basic question of this book is the same as that of Randall Collins's *Sociology of Philosophies* (Harvard University Press, 1998), Is intellectual history possible?

At the end of chapter 1 is the deep question, "How would we tell whether the Chicago school was a 'thing'?" By a thing he means a pattern that transforms various inputs into distinctive outputs; the transformations being autocorrelated over time. Things then are unities that repeatedly connect causes to distinctive effects. This seems to me an excellent definition of "thingness." The thing here is a pattern of conflict and cooperation among various tendencies (symbolic interactionism, urban sociology, surveys, and social organization) in the University of Chicago Department of Sociology, especially in the 1920s and 1930s. The most indicative output is dissertations that had a big influence on the discipline, though the person afterward had only marginal influence: a delicate adjustment of measure to concept. If students do better work when under the strong influence of the department than later, this shows that the causal factor is not the student, but instead the department. Abbott's conception of intellectual history is shaped by the question of whether such a jumble has distinctive effects: I doubt if many intellectual historians will follow him. According to his preface, his colleagues at Chicago have already chosen not to follow, which is a pity. Abbott's answer to the big question is that sometimes you can demonstrate that an intellectual movement has a structure that produces distinctive effects.

Abbott's notion is roughly that the clash and cooperation of different lineages in the department and in the *American Journal of Sociology* produced analysis of time-and-place-specific social structures. Time and place entered into the constitutions of the patterns of social life being studied, produced their thingness, and so produced the way they connected cause to the effect. The concentration on the significance of time and place in all the traditions then, he thinks, created the superiority of dissertations. An academic journal is, however, the "nonproductive" or

Permission to reprint a book review printed in this section may be obtained only from the book review's author.

"marketing" part of the production of knowledge and so requires a different history, more outwardly oriented. Abbott conceives of the history of *AJS* as becoming progressively a different thing than the department and so less tied to the "Chicago school." Albion Small started marketing sociology to reformers because he thought sociology and reform were different versions of the same thing. But the journal ended up marketing especially to sociologists not at Chicago, rather than to sympathetic ministers and social workers. From the 1920s to the 1950s, *AJS* marketed the wonderful intellectual variety of the Chicago department, and gradually that of Harvard and Columbia as well, to American Sociological Society members. Since the 1950s, it has marketed refereed scientific excellence, much as *American Sociological Review* has, though with a slightly more intellectual and historical cast and more toleration of ethnography, thus becoming a part of a different "thing," the discipline of sociology as refereed knowledge.

The market for refereed excellence seems to Abbott to be steadier than the market for Chicago departmental distinctness, though less backed by the production of exciting new knowledge. Abbott suggests that hardly anyone who buys the excellence actually reads it. The growth of refereeing work in the discipline as a whole, and what that growth implies for the decline in the advantage in competence of referees versus authors, deserves more serious study than Abbott gives it. Everyone who has had a referee get the argument of his or her paper directly backward has wondered about calling it "peer" review.

Abbott's last chapter is "presentist," oriented to what we should do with the Chicago tradition. He chooses as capstones the "pure structuralism" of Harrison White's network theories, or of his own analyses of the analogies between temporal sequences, as a way to turn out better dissertation work than students will ever do again.

I would choose a different capstone to imitate from the Chicago school: the complex project represented in Otis Dudley Duncan and others, *Metropolis and Region (Resources for the Future)*, 1960; Duncan, Ray Cuzort, and Beverly Duncan, *Statistical Geography* (Free Press, 1961); and Beverly Duncan and Stanley Lieberman, *Metropolis and Region in Transition* (Sage, 1970). This Duncan tradition brings together social organization, substantive network theory, urban sociology, and quantitative methods. The main methodological innovations are not for statistical inference or estimation, but for building measures to reflect the substance of structural forces on urban life. For instance, the network analyses are often of sizes of flows among cities and regions, whose varying size causes varying effects on city structures and locations, rather than links being merely present or absent as in White. The history in it is rates of growth rather than Abbott's sequences. Duncan students and postdocs of that period often wrote great dissertations, but they also kept on writing important books and articles. Beside these Chicago school masterpieces, White's and Abbott's network and sequence work look to me like "mere tinkering." I am not sure the Duncan sort of thingness is still in the department, nor in the journal.

Re-Orienting Western Feminisms: Women's Diversity in a Postcolonial World. By Chilla Bulbeck. New York: Cambridge University Press, 1997. Pp. xi+270. \$59.95 (cloth); \$19.95 (paper).

Lourdes Torres
DePaul University

Chilla Bulbeck challenges white middle-class feminists to move away from Eurocentrism and toward worldwide feminisms. In the service of this goal, Bulbeck leads her readers on a whirlwind tour of feminist thinking across the globe. The book departs from the critiques that women of color have been making of white Western feminisms for the last several decades. Bulbeck encourages white Western women to question their tendency to construct their identity in opposition to Third World women. In an attempt to challenge both simplistic dualism and universalism, the author not only presents a reading of non-Western women's thinking on an array of global issues, but also explores how these women perceive Western feminists. An important objective of this exercise is to find connections among women in situations of economic, social, and political inequality, rather than through assumptions about women's common essence. While *Re-Orienting Western Feminisms* rejects the homogenization of women's experiences, it argues that coalitions can be built around issues of similar concern to women across the globe.

After introducing her project and summarizing the tenets of Western feminisms, Bulbeck contrasts white Western feminist positions on a range of global issues with alternate analyses offered by women from African, Asian, and Latin American countries. Bulbeck skillfully brings together an enormous wealth of research by and about non-Western women through consideration of topics such as Western individualism versus community; the construction of mothers, wives, and sisters; the understanding of sexual identities; and the international traffic in women.

In the chapter on individualism versus community, which grounds all of the subsequent arguments in the text, Bulbeck argues that Western feminist philosophy posits an opposition between individual rights and community relations. White women construct themselves as autonomous agents with an awareness of rights; conversely, they construct women of color as subject to roles and obligations in their communities. According to this framework, white feminists are more advanced and enjoy more freedom than Third World women. Bulbeck argues that white feminists must reject this Western configuration of superiority and instead explore the tension between autonomy and relatedness. The author examines the complex ways that women worldwide negotiate the relationship between autonomy, rights, and obligations. She rejects the assumption held by some white feminists that women all over the world want to be more like Western women or that their ideal is to gain

the same rights as men. Bulbeck advocates the notion of "world-traveling" (proposed by Maria Lugones and Isabelle Gunning) or in other words, understanding unfamiliar practices in their cultural, social, and political context. Cultural practices like infibulation, dowry murders, and sati should be considered without prejudging them as barbaric or evading them by slipping into a mindless relativism. "World traveling" also entails looking at ourselves as others might see us; an especially useful tactic in this regard involves examining sites where analogous situations occur. For example, although Western culture has naturalized and normalized "cosmetic surgery," such as breast enlargements or liposuction, these practices may seem barbaric to women in other cultures. The connection could be made that women in the West undergo disfigurement under similar pressures that compel mothers in other cultures to demand clitoridectomies for their daughters to render them marriageable. The point is that women who partake in such practices should be understood within their historical and cultural context; they should not be viewed as ignorant victims needing to be rescued by Western feminists.

In her quest to place everything in its cultural context, Bulbeck sometimes undermines her own argument. In a discussion on the practice of dowry deaths in India, Bulbeck correctly points out that while Western feminists express outrage at the estimated 2,000 cases of dowry-related deaths that occur annually in the Indian home, they do not express the same horror at the approximately 2.5 million cases of robbery, rape, and assault that occur in U.S. homes. While this is a valid point, she then cites research suggesting that all incendiary deaths in the Indian context should not be attributed to dowry murders since Indian women occasionally set fire to themselves as a way of drawing attention to their problems. This attempt to minimize the number and significance of dowry-related deaths in Indian society diminishes her argument.

While Bulbeck documents differences between women in terms of worldviews and priorities throughout the book, she concludes that coalitions among women worldwide are possible if they are built on mutual respect and an understanding both of the dynamics of the relevant issues, and also one's own biases and ignorance. Successful coalition politics are built on a thoughtful understanding of one's own positionality as well as the perspective of those with whom one is working. Above all, in order to enter into coalitions, Western women must refrain from seeing only victimization when they examine the situations of non-Western women, and they must concurrently analyze their own unearned positions of privilege and power in relation to other women.

Gender Violence in Africa: African Women's Responses. By December Green. New York: St. Martin's Press, 1999. Pp. 298. \$49.95.

Jacqueline Bhabha
University of Chicago

Why has postapartheid South Africa witnessed a dramatic explosion in the incidence of rape, such that on current trends one in three South African women are expected to be raped in their lifetime? Why is there an almost perfectly inverse relationship between genital cutting and wife beating in African countries—where genital cutting is absent wife beating is endemic and vice versa? Why are witchcraft accusations suddenly on the increase in a range of different African countries? Answers to these and a range of other fascinating questions are proposed in December Green's ambitious and richly researched account of gender violence in Africa. Using feminist theories on postcolonial gender relations to frame a "regional case-study" of gender violence in 47 countries in sub-Saharan Africa, Green draws together sociological and anthropological literature documenting a broad range of phenomena, from traditional practices in respect of bridewealth to international responses to mass rape.

The central claim of the book is familiar. Gender violence in Africa, as elsewhere, is not the result of aberrant individual behavior or domestic discord, but a central aspect of the social control of women. It is part of the "politics of everyday life." A subsidiary claim is also uncontroversial: though gender violence is globally endemic, its particular manifestation reflects socially determined options: "It erupts at the intersection of social forces and individual choices" (p. 6). Thus in some countries, wife beating is considered socially acceptable, a common method for settling domestic disputes: in Tanzania, for example, a government survey reported that 90% of women claimed to have been battered (p. 41); in Zambia, "a man could kill his wife and incur less of a jail sentence than if he stole a goat" (p. 41). On the other hand, in some African states, such as Malawi, Côte D'Ivoire, Togo, and Somalia, wife beating is considered socially unacceptable, and as a result, incidents are rare and further reduced by enlightened political leadership on the topic. A similar point is made for "genital cutting," the non-value-laden term for female genital mutilation. Its prevalence reflects traditional support for the practice, a manifestation in those countries of the high premium placed on female chastity. But, Green argues, sustained or even increasing rates of FGM might also indicate a community's choice of survival mechanism in the face of brutal political or economic upheaval, a nostalgia for tradition, a quest for solace in stability or communal continuity. In Sierra Leone, for example, mass circumcisions were held to celebrate the return of displaced persons at the end of the war and to mark the longing for some form of normalcy (p. 51).

Rather than organizing her material in terms of country or violation,

Green divides her descriptive account into three overarching categories of relevant "power relations": sexual, economic, and state. The classification mirrors that adopted by the UN Special Rapporteur on Violence against Women in her recent reports. This approach has the merit of highlighting the links between gender violence across regions and social groups and privileging a broad explanatory framework, including the destructive effects on much of Africa of aspects of the contemporary global economy such as transnational migration and structural adjustment policies. Green is to be congratulated for tirelessly presenting the relevant theoretical arguments as she describes her material. However, the text sometimes reads like a faithful précis of research reports—where divergent material and argument are summarized sequentially rather than being integrated into a coherent argument. Thus, on the one hand we learn that uprooted, socially disconnected women are particularly vulnerable to sexual violence, including rape: Somali women in Kenyan refugee camps are a well-known example cited (pp. 91–92). On the other hand, it is claimed, with some force, that women who are independent, seen to be outside the sexual control of their male peers, whether because they are commercially successful or elderly, are particularly vulnerable to allegations of witchcraft and gender-determined killings (p. 81). How are these discrepant facts to be reconciled?

The final third of the book deals with the topic of the subtitle—African women's responses. It includes a fascinating account of what Green rather counterintuitively calls the "politics of disengagement," namely, traditional, customary, and community-based responses by women to their oppression. Though a lot of the material is presented with no indication of whether these practices are contemporary or widespread, they complement the more familiar account of NGO-based activism in the final chapter. Green argues that by using traditional resources to subvert the power of men over their lives—be they mechanisms of individual withdrawal or collective action—women have managed to garner an effective set of strategies independent of state action. The account is richly suggestive and deserves a wide audience—not so much, as Green claims, to dispel (unattributed) notions regarding the alleged passivity of African women, but rather as a testimony to the creativity and determination of this population of survivors. Predictably, Green claims that NGOs that base their actions on eclectic strategies that draw together traditional and contemporary institutions have the best chance of success. Examples include finding alternative work for excircumcisers, replacing cutting with alternative coming-of-age ceremonies that incorporate traditional teaching and celebrations, and building on traditional institutions such as women's secret societies to seek changes in political consciousness.

This inspiring book is an interesting combination of a scholarly textbook and a political commentary. It will be invaluable for students of gender relations in Africa, providing a well-organized, if somewhat dense, compendium of research on the topic. Occasionally the author lapses into

generalized critiques of "Western feminism" and celebrations of "African women," which belie the sophistication of much of her nonessentialist, detailed analysis of this disturbing topic.

Asian/American: Historical Crossings of a Racial Frontier. By David Palumbo-Liu. Stanford, Calif.: Stanford University Press, 1999. Pp. vi+504. \$65.00 (cloth); \$24.95 (paper).

Yen Le Espiritu
University of California, San Diego

In his influential writings on urban life, eminent Chicago sociologist Robert Park argues that the modern age is marked by a new global economy and dramatically increased migrancy. According to Park, the global economy has produced one of the fundamental problems plaguing the modern state: the accommodation of migrancy, race, and ethnicity. Drawing on and extending Park's claim that the Pacific represented a formidable challenge to modern America in terms of both race and culture, David Palumbo-Liu's *Asian/American* argues that the historical formation and imagining of America in the 20th century has had everything to do with Asia—or more specifically, with "westward expansion" across the "Pacific frontier" and the movement of Asians onto American soil. His massive inquiry spanned from the importation of Asian labor in the mid-19th century, to the territorialization of Hawaii and the Philippines in the late 19th century, to the wars with Japan, Korea, and Vietnam and the Cold War with China, and to today's Asian Pacific Economic Cooperation group. Each historical juncture, he argues, contains a particular conception—as well as a contingent fusion—of "Asian" and "American."

In this dense and wide-ranging inquiry into questions of Asian and American identities, Palumbo-Liu skillfully demonstrates that these historical crossings were bidirectional: as America crossed over to Asia, Asians came to the United States. Thus one of the fundamental questions for modern America has been how to grapple with the presence of Asians already in America *and* with the imperative to develop its neocolonial interests in Asia. Drawing on an impressive range of scholarship and materials—sociological, historical, cultural, medical, anthropological, geographic, economic, and political—Palumbo-Liu reads the predication of Asian to American in and through the discourse of body, psyche, and space. He contends that the historical and continued racialization of Asian America draws on particular imaginings of the Asian/American body and psyche, and the ways Asian Americans might or should occupy a particular place/space in America.

Echoing other histories of Asian America, Palumbo-Liu shows—through insightful analyses of immigration laws, economic and social pol-

icies, and cultural practices—that Asian Americans have been marginalized, excluded, or relegated to particularly constrained roles in the American imaginary. However, in a productive turn, he asserts that Asian American history has not simply been a matter of constant degradation and marginalization. Calling attention to the pervasive “model minority” myth, he argues instead that it is the specific nature of Asians in America and in Asia to be held up as models at various points in history, not only for racial minorities to emulate, but for whites as well. Ever attentive to historical contexts, Palumbo-Liu situates the emergence of the model minority myth within the contexts of domestic civil rights struggles and Asian growing influence in the global economy. It is this anxiety over America’s seemingly weakened position at home and globally that led to restorative strategies deployed to salvage and distinguish America from Asia. Palumbo-Liu argues that these strategies—which simultaneously emulated and reparticularized Asians—ultimately produced an Asian America that existed in a liminal state. As an example, the “foreignization” of successful, well-to-do Asian American contributors to the 1996 presidential campaign attests to the reality of this *in transit* state. Through nuanced analyses of films, novels, and other cultural “texts,” Palumbo-Liu goes on to show how Asian Americans have both inhabited and resisted these roles and categorizations, inventing their own particular subjectivities as Americans, changing and (re)shaping America in the process.

Sociologists, especially those who work in the areas of race and ethnicity, would be most interested in Palumbo-Liu’s extended and oft-critical engagement with Robert Park’s writings. Through close readings of Park’s works, Palumbo-Liu contends that Park’s conceptualization of race—as a problem carried over from the premodern age—abstracts and brackets the issue of race and ignores the socioeconomic apparatuses that perpetuate and manage racism. According to Palumbo-Liu, Park’s abstraction of race threatens to displace marginality into the psyches and onto the body of racialized peoples. Drawing on the lives of Asian Americans, he calls attention to the perpetuation of the racial divide in the United States and argues instead for a materialist reading of race—one that is deeply rooted in national and global economic change.

Whether or not we agree with Palumbo-Liu’s contention that the formation of “modern America” is deeply attached to the Pacific region, we have to marvel at the ease with which he moves across disciplinary and area boundaries. I suspect many sociologists will feel lost—and might even get frustrated—as they wade through the oft-cumbersome language of cultural studies in the text. On the other hand, as a sociologist who believes in and strives to do interdisciplinary scholarship, I was impressed at how adept Palumbo-Liu was at handling sociological materials. It should be incumbent upon us sociologists to do the same—to read and write across disciplines.

Global Transformations: Politics, Economics and Culture. By David Held, Anthony McGrew, David Goldblatt, and Jonathan Perraton. Stanford, Calif.: Stanford University Press, 1999. Pp. 538. \$75.00 (cloth); \$29.95 (paper).

Giovanni Arrighi
Johns Hopkins University

This imposing addition to the proliferating literature on "globalization" aims at developing an account of the phenomenon that is both historically grounded and informed by a rigorous analytical framework. Like many before them, the authors define globalization as consisting of those processes that have transformed the organization of human affairs by generating transcontinental and interregional flows and networks of activity. But more explicitly and rigorously than in the existing literature, they conceive of these processes as occurring along four distinct spatial-temporal dimensions: what they call "extensity" or stretching; intensity or regularity; velocity or speed; and impact or degree of enmeshment of the local and the global. Four different types of globalization follow logically from this differentiation of dimensions: (1) *thick globalization* (high extensity, high intensity, high velocity, high impact), (2) *diffused globalization* (high extensity, high intensity, high velocity, and low impact), (3) *expansive globalization* (high extensity, low intensity, low velocity, and high impact), (4) *thin globalization* (high extensity, low intensity, low velocity, low impact).

Each of the eight central chapters of the book attempts to gauge the extent, intensity, velocity, and impact of globalization in successive epochs of world history (premodern, early modern, industrial, contemporary) by analyzing the evidence from a particular angle of vision: (1) state making, (2) war making, (3) trade, (4) finance, (5) production, (6) migration, (7) culture, and (8) the environment. The concluding chapter brings together the findings of these partial accounts into an overall assessment of the novelties of contemporary globalization in comparison with earlier forms of globalization. In terms of the typology sketched in the introduction, the main conclusion is that globalization has a long history and that the central fact of this history is the epochal transformation of the "thin" globalization of premodern times into the "thick" globalization of the contemporary period. Surprisingly, nothing is said explicitly in the concluding chapter (nor indeed in any of the preceding eight chapters) about the historical significance of the other two types of globalization ("diffused" and "expansive"). Nevertheless, in introducing the typology, the authors claim that the early modern period of Western imperial expansion comes closest to the "expansive" type of globalization and that the "diffused" type "has no historical equivalents." It is "a state of affairs which, normatively speaking, many of those critical of the excesses of contemporary economic globalization might find desirable" (p. 22).

In short, the book describes a trajectory whereby a globalization that had been "thin" for centuries or even millennia in premodern times became "extensive" in early modern times and "thick" in the contemporary

period. The future of the trajectory is left open, but the authors seem to join those who advocate the transformation of today's globalization from "thick" to "diffused." This open-ended conclusion is as welcome as the book's injection of a badly needed dose of historicity into the globalization debate. There are nonetheless two main problems both with the conclusion and the analysis on which it is based.

First, the authors' difficulties in classifying late 19th- and early 20th-century globalization betrays the fact that they have not resolved one of their central issues, that is, the dispute on whether the thickness of contemporary globalization is as unprecedented (and constraining of state action) as the "hyperglobalizers" maintain, or whether it is comparable to the thickness of globalization in the late 19th- and early 20th-centuries, as the "skeptics" maintain. The issue is important because the processes of globalization of a century ago were reversed catastrophically during and after World War I. Our assessment of the irreversibility of contemporary globalization thus depends on whether we think that globalization at the end of the century is qualitatively different from what it was at the beginning. The problem is that the evidence presented in the book is inconclusive. Most of its analytical chapters suggest that today's globalization appears "thicker" than it was a century ago but not so much "thicker" as to deserve being classified as a distinct type of globalization. Moreover, at least one chapter (that on migration) suggests that today's globalization is actually "thinner" than a century ago. This kind of inconclusive and contradictory evidence can be expected to leave hyperglobalizers and skeptics as convinced as ever of the validity of their respective positions.

Closely related to the above, and more serious, is the second problem, that is, the book's failure to provide any coherent explanation (or set of explanations) of the epochal transformation of globalization from "thin" to "thick," which it describes. No explicitly stated theory or hypothesis guides the selection and presentation of the empirical and historical material. Contrary to the authors' claim (p. 25), their typology of globalization is not "a theoretical model." It is a purely formal classificatory scheme. Useful as it is in guiding the observation and measurement of processes of globalization, it does not tell us anything substantive about those processes. As a result, the reader is taken through a prodigious amount of observations and measurements, the meaning and significance of which often remain unclear. In view of the immense empirical scope of the book, that is really unfortunate.

The Discipline of Teamwork: Participation and Concertive Control. By James R. Barker. Thousand Oaks, Calif.: Sage, 1999. Pp. xiv + 207. \$29.95.

Randy Hodson
Ohio State University

Teams are the cutting edge of organizational practice and have become an important focus of scholarly analysis. James Barker has succeeded in

writing a clear and definitive statement on the costs of participation for workers resulting from heightened peer scrutiny and work intensification in team settings. His work offers important insights on the subtle influences of teams in remaking the experience of work in contemporary society.

Barker coins the term "concertive control" to describe the self-disciplining nature of teams. The term denotes workers acting *in concert* with each other to control their own behavior. Concertive control is the successor to previous forms of control based on direct supervision, technology, or bureaucracy. Barker argues that this new form of control is the tightest and most effective yet.

The setting for Barker's study is an electronic circuit board manufacturer (ISE) in Colorado's growth corridor stretching from Denver to Colorado Springs. The core methodology is classic ethnography based on observation and interviews. To this Barker adds a rigorous analysis of the rhetoric used by team members in the process of developing new work norms and disciplining team members. Through these methods, Barker is able to reveal the subtle world of group pressure that defines team production.

Teams at ISE operate within a world ultimately framed by economic coercion. The workers need their jobs. Its large corporate owner recently sold the enterprise to new owners, and fierce competition in the industry suggests that failure to innovate will result in bankruptcy. The values the teams seek to maximize are thus ultimately those of the corporation and marketplace: productivity and cost reduction. Barker's study is an examination of how these corporate values *become* the teams' values.

The core chapters describe the manner in which a team production culture emerges, is stabilized, and is maintained. The initial impetus comes from the vision of the inspired vice president for manufacturing. The stages of maturation include a reversion by teams to bureaucratic rules, sometimes even in excess of the past. The system ultimately succeeds through workers vesting their identities in a new community at work in which they are now active participants rather than the pawns of more powerful supervisors.

Teams, however, are often harsh taskmasters. New recruits who do not fit in quickly "find the door." As Barker notes, "Before the change to teams, the line supervisors would generally tolerate some degree of slackness. . . . But now the team members exercise their newfound authority with much less patience" (p. 80). Team production thus includes substantial costs for workers. These costs include more time and energy spent on the job and more stress associated with the hypersupervision of peers acting as supervisors.

Barker's critique of team production is that it is part of an increasingly limiting and constraining organization of work life. The image here is almost that of an omnipresent Big Brother—but the operational mechanism is peers and peer pressure, not cameras or spies. In the final chapter, Barker offers a list of strategies for ensuring that the productivity gains of teams for corporations are paralleled by gains for workers. These strategies mandate that teams be open to values more humane and inclusive

than economic survival and profit alone. Most centrally, Barker's suggestions involve open discussion and dialogue on a full range of topics, unbounded by corporate agendas. These suggestions will, hopefully, provide a blueprint for widening the values to be pursued in production teams and, ultimately, for humanizing team production.

A possible limiting factor in Barker's analysis (and suggestions) is that he does not develop the benefits of team production for workers with the same rigor with which he analyzes its costs. Repeatedly, workers say that they "love this place" and that they would never go back to the previous management system. We cannot just dismiss their enthusiasm as false consciousness.

Why do workers like teams in spite of more work, more effort, and an intensification of supervision? Elements of an answer are hinted at in Barker's work. Work takes on new and important meanings for workers as they find themselves active participants in the creation of their own work experiences. Workers gain an increased sense of identity and community in workplaces involving active participation. Other researchers have noted increased meaning and identity in contemporary work settings as well (see Arlie Hochschild, *The Time Bind* [Metropolitan Books, 1997]). Increases in interpersonal and communication skills have also been noted as an important side gain for workers in participatory settings (see Vicki Smith, "Employee Involvement, Involved Employees: Participative Work Arrangements in a White-Collar Service Occupation" [*Social Problems* 1996:166-79]). Increased productivity and profits resulting from teams also suggest that team workers are more valuable (and perhaps less replaceable) than workers under more rationalized bureaucratic systems. Had Barker developed a more complete vision of the potential benefits of team production for workers, the utility of his suggestions for widening the gains of team production might have been further increased.

The Discipline of Teamwork is essential reading for researchers and graduate students in organizational behavior, the sociology of work, and industrial relations. The analysis of subtle aspects of team relations is the best I have read to date. If the next round of studies highlights gains as well as costs and is as rigorous and insightful as Barker's, we will be well on our way to achieving continuous improvement in organizational studies with James Baker as our team facilitator.

Taking Care of Business: Samuel Gompers, George Meany, Lane Kirkland, and the Tragedy of American Labor. By Paul Buhle. New York: Monthly Review Press, 1999. Pp. ix+315. \$18.00 (paper).

Kim Scipes
University of Illinois, Chicago

Paul Buhle has written an important history of the U.S. labor movement, one that tries to understand why this labor movement has degenerated

to the extent it has: the unionization rate in the private sector, once almost 35% in the early 1950s, is back to the same level (under 10%) it was at the time of the Great Depression. While not the first to note its internal disintegration, he does not attribute its attenuation to any "iron law," but rather to the historical development of bureaucracy within the top levels of the union movement and the conscious choices made by bureaucratic leaders, often at the expense of the majority of workers' at home and abroad.

Buhle focuses on top-level leadership in the American Federation of Labor (AFL) and, later, the American Federation of Labor-Congress of Industrial Organizations (AFL-CIO). He concentrates on three of the most important leaders: Samuel Gompers (president, 1886-94 and 1895-1924), George Meany (1952-79), and Lane Kirkland (1979-95). Clear on the problem, labor bureaucracy, Buhle ties its development into the larger context: "America's historic labor bureaucracy is . . . finally and without doubt connected to the creation of empire" (p. 15).

This is a provocative thesis that he argues well. He claims that the structural sources for the continued betrayal by leaders are "a share of the promised imperial benefits and . . . the shared psychological satisfactions of a personal superiority of the white 'aristocrat of labor' over all the lowly, whether 'foreign,' female or non-white" (p. 19). Starting with a quick history of the "American" (in reality, white male) working class, he argues that, in the face of these workers' degradation by capital in the 1800s, skilled workers responded by allying with capital against nonwhites, nonmales, and non-"Americans" in efforts to try to retain their respective positions and related status. He also argues that Gompers, Meany, and Kirkland, respectively, albeit each in their own particular way, consistently acted to further these same interests over time.

Buhle marshals an impressive array of evidence to support his thesis. In addition, he is particularly strong on understanding developments among Jewish workers and their unions, as well within the labor movement in and around New York City. Particularly during the regimes headed by Meany and then Kirkland, he has a strong understanding of top-level labor movement culture.

Buhle's book will be of interest to sociologists who focus on organizations and those interested in labor. His is a welcome corrective, as he recognizes the importance of the labor movement and its internal organization, subjects little studied in recent times, as most recent attention has been directed toward organization of corporations and business in general.

However, Buhle's view is ultimately restricted. He places the onus for labor's historical failure on the leaders within the organization, but he fails to consider why rank and file union members have continued to elect and then follow these leaders, even as the disintegration of the labor movement has become more obvious since the late 1970s.

I think this limitation flows from his approach to the subject. Although

suggesting he is doing an organizational history of the labor movement, he is, in reality, doing a cultural history of the leadership. To put it another way: he focuses on the culture developed by the top leaders and, rather than using it to understand the whole, he (in effect) argues that it is the whole. Rather than a sociological study of the organization—of which this culture would be an important component—he puts forth a cultural study of one component as though it were the whole. Finally, while Buhle's concentration on the labor bureaucracy is an important part of the story, I think he probably would have better served his purposes by focusing on struggles for democracy and inclusionary types of unionism in the face of dominating efforts by labor bureaucrats.

There is no question that, in the majority of situations, the bureaucrats have prevailed. But approaching things from the bottom-up rather than from the top-down would have acknowledged struggles for democracy and social justice that have come out of the labor movement and would have allowed recognition of the times they have succeeded in the face of extreme opposition. Because of Buhle's focus on the labor bureaucrats, he must go outside the logic of his argument to explain emerging opposition to business unionism by rank-and-file activists and, ultimately, the "earth-shaking" repudiation of Lane Kirkland in 1995 and his subsequent replacement with John Sweeney.

This is a very strong effort, and it deserves wide attention. There is a lot of material to be mined here, but there still remains much more to be understood about the labor movement during the times of Gompers, Meany, and Kirkland.

Rethinking the Labor Process. Edited by Mark Wardell, Thomas L. Steiger, and Peter Meiksins. Albany: State University of New York Press, 1999. Pp. x+279.

Martin Parker
University of Keele

Twenty-five years after Harry Braverman's *Labor and Monopoly Capital* (Monthly Review Press, 1974), research and controversy within labor process (LP) treatments of contemporary work organizations continues unabated. What is unusual about this collection is that it originates in the United States, when the majority of the work coming out from the British LP conferences over the last 15 years has been Western European. We begin with an introduction by Wardell, which sets up some of the central ideas and problems with an admirable clarity. Kraft contributes a provocative chapter on the limits of Taylorism and then move toward total quality and business process reengineering. For Kraft, the distinction between conception and execution so central to Braverman's analysis has given way to a globalized form of adding value based on the distinction

between routine and exceptional labor. The next chapter, by Burris, suggests that what she calls "technocracy" has resulted in a similar distinction, between the privileged experts (with their associated knowledge claims) and the nonexperts who are increasingly routinized.

Rogers follows with a debunking of some of the more optimistic claims made about temporary workers and their place in flexible labor markets. These temporaries are not "symbolic analysts" with "portfolio careers," but an overqualified and underpaid reserve army of labor. Greenbaum makes similar points but ties them to an argument about technology's ability to disentangle work from fixed times and places. Since rationalizing functions are now built into the technical infrastructure, they can increasingly be outsourced on a global level. We then move to two chapters that are both intended to serve as correctives to Braverman's determinist and productivist view of history. Gartman's chapter identifies four periods within the U.S. auto industry and relates them to the rise of consumerism and crises of accumulation. Isaac and Christiansen follow this with a detailed analysis of the Lordstown General Motors strike (mentioned by Braverman), which redescribes it as an outcome of particularities, not abstract general tendencies. Then, come two chapters with a more quantitative emphasis. James and Laura Geschwender discuss occupational sex segregation and insist that the social division of labor is as important as the technical, while Steiger's chapter contains a neat statistical demonstration of the continuing relevance of the labor theory of value. The final contribution, by Smith and Thompson (the only U.K. authors) reevaluates the various ways in which LP has been developed, with their usual sideswipes at Foucauldian revisionism set within a general skepticism about claims of postindustrial paradigm change.

So, what to make of a collection like this? The first issue seems to be whether LP can stand the definitional expansion and revision in which many of these chapters engage. Whether this is "neo-Marxism" or "post-Marxism" might be one way of putting the question. Some of these authors seem to borrow as much from Daniel Bell as from Braverman; others make class into a variable alongside gender and race; and yet others deny Braverman's production focus, or conception of resistance, skill, management strategy, and so on. In general, there seems to be a move toward a more contingent, rather than determinist, analysis. Braverman's assumption that "technology equals deskilling equals proletarianisation" is revised or denied in all the work in this book. So what is LP then? On the one hand, we have Wardell—"A labor process approach encompasses a broad integrative view that draws liberally from a variety of disciplinary backgrounds" (p. 15). On the other hand, we have Smith and Thompson—"Its core theory *merely* recognises that competitive relations compel capital to constantly revolutionise the labor process and that within that framework, capital and labor will contest the character and consequences of such transformation" (p. 211; emphasis added). As I understand them, neither of these definitions would necessarily exclude (for example) Weber, Foucault, or Castells. This may not be a problem, but

it does make me wonder how much definitional coherence a term needs to be useful. Is LP like "postmodernism," a word that means whatever you want it to?

Second, for this English reviewer, the book shows that there are some important differences of emphasis between U.S. "labor" process theory and U.K. "labour" process theory. Interestingly, nine of the eleven U.S. authors write from sociology departments. The two U.K. authors are based in management—as are most of the other contributors to the LP conferences in the United Kingdom. As Wardell hints in his introduction, British sociology of work in the 1980s was far more steeped in Weberian and Marxist traditions. As a result, Braverman stimulated more scholarship in Western Europe than in North America. That most U.K. writers now work in a business school context is rather ironic, but it also reflects the lack of interest in work organizations within contemporary U.K. sociology. While this seems (on the strength of this collection) not to be true of U.S. sociology, the latter contains quantitative methods and postindustrial assumptions that are fairly alien to the former and perhaps to Braverman's original text as well.

Then again, neither of these observations are disabling to those who continue practicing LP analysis, and this book is a useful survey of many of these ideas. From my point of view, critical treatments of work organizations, wherever they come from and whatever they are labeled, are a valuable corrective to the deification of "management" and to accounts of capitalism as the end of history. If labor (or "labour") process theory keeps alive the possibility of things being different, then it will deserve continued interest.

Dragon in a Three-Piece Suit: The Emergence of Capitalism in China.
By Douglas Guthrie. Princeton, N.J.: Princeton University Press, 1999.
Pp. xiv+302. \$39.50.

Ching Kwan Lee
University of Michigan

Sociological studies of China's market reforms have in recent years generated spirited debates, thanks partly to researchers' self-conscious application of sociological theories to Chinese materials and partly to the insights Chinese research lends to the field of comparative transition. But as Anthony Oberschall (*American Journal of Sociology* 101 [4]: 1028–41) has rightly warned us, there is danger in the predominant trend of "statisticism": a reliance on large-scale surveys that accumulate more data than knowledge. Even the most sophisticated statistical analyses of individual and household income or aggregate indicators of productivity may not be effective tools for us to grapple with "transition" as the mutual determination of institutional power and practices of social

actors. Doug Guthrie recognizes this important lacuna, and his book makes a highly commendable attempt to go beyond the literature. Zeroing in on the firm level of analysis and aiming at discovering practice rather than policy, Guthrie deftly weaves together intriguing in-depth interview data and systematic survey analyses, fleshing out their implications for organizational theories. His main question is, What accounts for a rationalizing process of Chinese firms that have been adopting a number of strategies resembling those in advanced market societies? He has marshaled an impressive pool of firsthand data in Shanghai, including survey responses from 81 medium and large state-owned enterprises, randomly selected from four industrial sectors, and 155 interviews with industrial managers, legal professionals, and local officials. Guthrie attempts with admirable clarity to adjudicate the relative merits of rationalist/efficiency perspective and institutional theory of organizational behavior, in the light of Chinese data. Core chapters of the book discuss in detail these firm-level transformations, including formalization of intraorganizational labor arrangements like wage, nonwage benefits, and labor contracts; market strategies of price setting and service-sector diversification; adoption of the company law and thereby a new governance structure; and interfirm negotiations regarding clauses in joint-venture contracts.

His general finding is that organizational change varies among Chinese companies depending on their location in the hierarchy of state administration, the existence or otherwise of formal joint-venture relationships with foreign partners, and the educational background of general managers. More specifically, firms in the upper levels of the administrative hierarchy, or those that used to be the most protected in the planned economy, are experiencing greater financial and organizational uncertainty than their lower-level counterparts. Once reform sets them adrift, upper-level firms are more prone to rationalization. Likewise, firms having extended contact with Western management practices, via formally trained managers or joint-venture contracts, will be more likely to incorporate Western practices as a symbolic exercise or a strategy to appeal to foreign partners. Guthrie's theoretical point is that rather than motivated by a rational calculus of profit maximization, as the rational/efficiency theory suggests, firms' espousal of organizational change is a mimicry of organizational models widely perceived as successful. The search among Chinese firms, he argues, is for legitimacy rather than efficiency, a conclusion that vindicates the claims of the institutional perspective.

As all pioneering works, this ambitious book opens up interesting questions that need further exploration. Despite his insistence on the need to distinguish, on the one hand, central government policies and performance indicators and, on the other, decision, practices, and strategies of economic actors, I still would have liked to have seen more data on how firms actually operate. His evidence tell us which firms would set up what organizational structure, but he has not delved deep enough into the mo-

tivation of actors and the ways these structures affect actual operation of management, both of which have implications for his discussion of "rationalization." Some of the interview quotes contain extremely revealing accounts, which Guthrie seems to have let pass too easily. For instance, in chapter 6, his interviewees consistently and explicitly reminded him that the situation is very "chaotic" (p. 132), with the same organizational features being put to different uses, some more rational than the others. Thus, when enterprises apply to become companies, widely different organizational consequences follow. One may also wonder whether Chinese firms confront problems of conflicting interests, intraorganizational power dynamics, and contending ideas and models of change in this period of momentous transformation. Notions of mimicry and isomorphism, together with their underlying assumption of cultural diffusion and learning, may have concealed these central dynamics of firm restructuring.

Another unfortunate weakness of the study is that Guthrie starts his institutional analysis by imposing models from the West onto Chinese reality, instead of constructing institutional models from grounded data. Due to an overriding concern to establish the distance between Chinese firms' reality and Western firms' model, Guthrie has forsaken an opportunity to analyze fascinating patterns of organizational practices that he claims he aspires to achieve (p. 11). Reading the quotes on pages 115 and 188, for instance, we can see informants differentiate the varying situations under which they would negotiate prices or make use of *guanxi* to accomplish certain types of transaction. But when Guthrie compresses these organizational strategies into dummy variables (presence or absence of price negotiations, importance or unimportance of *guanxi*) for the statistical manipulation and testing of theories, he buries the complex but nonrandom logics of managerial practices his informants are trying to explain. An alternative approach would be to uncover the patterns whereby different kinds *guanxi* may be mobilized to achieve diverse goals. After all, many researchers would doubt the validity and utility of a conclusion as general and sweeping as "the declining significance of *guanxi*." Finally, it is very obvious that Guthrie has expounded an American perspective on Chinese reforms, as reflected in his choice of what constitutes a "rational" organizational model and in his discussion of practical implications for American trade and human rights policies in the concluding chapter. One might also note, in this regard, the orientalist title of the book. My concern is not so much with the lack of postcolonial sensibility per se, but more with its implications for the goals Guthrie sets for himself. Would this American centeredness have blinded him to the fact that Chinese managers are looking beyond the United States, say at Japan, Korea, and Germany, for models of "rational" organization? All said, there is no denying that the labor this book embodies is immense. It is an original, timely, refreshing, and provocative volume, which will certainly galvanize the interest of a wide spectrum of readers in the scholarly, business, and policy-making communities.

Guests and Aliens. By Saskia Sassen. New York: New Press, 1999. Pp. xxi+202. \$25.00.

Charles Tilly
Columbia University

Spurred by Eric Hobsbawm, in 1996, Saskia Sassen published her *Migranten, Siedler, Flüchtlinge* as one of Germany's widely circulating Fischer Taschenbücher. Now the nonprofit and resolutely nonconformist New Press has published an English version of that historical essay, with a new preface addressed to North American and Western European audiences.

Rather than a factual survey, a theoretical synthesis, or a portrayal of migration's human dramas, *Guests and Aliens* arrays highly selective information about European migration streams since 1800 as a critique of common misconceptions and misrepresentations concerning today's immigration. Sassen writes vigorously, vindicating her breathless flight across so much history and geography with the hope "to widen the options we envision for dealing with immigrants and refugees by making a broader interpretation of why these people in motion exist in the first place" (p. 2). Against fears of mass invasion by threatening aliens, she argues strenuously (and correctly) that emigration generally involves only small shares of sending populations, that immigrants almost always constitute small minorities of receiving populations, that Europeans (and by extension, North Americans) have repeatedly exaggerated racial and cultural differences between immigrants and receiving populations, that most migration streams include substantial return migration, that some immigrants nevertheless settle permanently at the destination, and that since World War II (partly as a result of tightened legal restrictions rather than basic changes in migration dynamics) illegal migration has increased as a proportion of the whole.

Roughly half the book reviews 19th-century history. That half relies heavily on three outstanding European doctoral dissertations: Jan Lucassen's reconstruction of Western European migration systems around 1800, Abel Châtelain's study of temporary migration in France during the 18th and 19th centuries, and Klaus Bade's treatment of German labor migration between 1880 and World War I. (Even in so cursory a discussion, it is disconcerting to find no reference to the relevant investigations and overviews of Leslie Page Moch, Lynn Lees, Lawrence Schofer, Rudolf Braun, William McNeill, and Jan Lucassen's brother Leo.)

Sassen concludes her book's first half with a comparison of Germany, France, and Italy. The comparison emphasizes how small numbers of immigrants (in the German case, Polish immigrants, who brought out a nationalist streak in so cosmopolitan a figure as Max Weber) can generate large political panics. It also establishes how even rights-oriented regimes such as France's made exceptions for the exploitation of immigrants and how selective outmigration was even in such a great exporter of labor as Italy. A 20-page chapter takes us from the early 20th century to the eve

of World War II, arguing (again correctly) that states became much more heavily involved in controlling migration during this period and that one consequence of increased regulation was, paradoxically, the expansion of a distinctive category: the political refugee.

Leaping over the vast displacements that occurred during World War II, the book's last third follows a rapid survey of migration streams and their consequences since 1945 with a survey-cum-critique of contemporary European migration policies. Here Sassen turns to advocacy even more energetically than in the earlier chapters' side comments and summaries. She makes clear her cases for checking political discourse that raises the specter of mass invasion, for welcoming international standards and institutions to protect migrants' rights, and for moving European politics toward cultural and religious diversity.

For all my own sympathy with these positions, I wish Sassen had spent more effort identifying and countering the obvious objections: that cheap and easy labor mobility undercuts labor's political and economic strength, that disposable labor discourages capitalists from investing in widely shared productivity increases, that immigrant clusters encourage a politics of ethnicity, that remittances drain capital from national economies, that the European Union sacrifices local needs to general principles, and so on. I also wish someone had taken more editorial care with the errors that pock the book's text. Sassen generously declares that "all mistakes are mine" (p. xii), but one must wonder how two sets of editors, one of them German, could have let pass such howlers as the invention of a *Wischenschaftszentrum* (read *Wissenschaftszentrum*) Berlin (p. xi) and the claim (p. 37) that the North German Confederation sent 12 million men against France in 1870–71. (Bismarck's actual assembly of about half a million troops was, after all, already a tour de force.) Two of Sassen's major authorities, Nicole Fouché and Abel Châtelain, invariably appear in the text stripped of their accents. Nor is it true (p. xi) that Châtelain's thesis was "originally published in the 1930s"; indeed, Châtelain did not live to defend the dissertation or to see its publication in 1976. So important an argument as that of *Guests and Aliens* deserves more scrupulous documentation.

Global Futures: Migration, Environment and Globalization. Edited by Avtar Brah, Mary J. Hickman, and Máirtín Mac an Ghail. New York: St. Martin's Press, 1999. Pp. xi+243. \$69.95.

Peter J. Taylor
Loughborough University

The publishers have done very well here to get such important keywords into a reasonably short title. Good for search engines (and for easy book ordering by university libraries), but is the product good for readers? We are promised, rather extravagantly as it turns out, "a critical assessment of

the 'globalization thesis' through sustained analysis of the nexus of processes underlying social and cultural relations" (inside front flap of cover). In fact, what we have is one of three edited volumes derived from the 1996 British Sociological Association's annual conference whose theme was "Worlds of the Future." Everybody (apart from, presumably, those working in the marketing departments of publishers) knows that edited books are most unlikely to achieve a "sustained analysis," and this volume is no exception. Like most others of its ilk, it is variable in quality, and the contents are not as integrated as the editors would have us believe.

There are 11 chapters organized into four parts. The first part, "Imagining the Global," consists of two chapters, an introduction from the editors, and a chapter by Doreen Massey summarizing her ideas on power-geometries and time-space. The former consists of just four pages telling us how contested a concept globalization is, followed by 21 pages summarizing the remaining chapters. Massey's "imagining globalization" does a much better job at introducing the subject by identifying three approaches to the relation between space and society: through a space of enclosed places, a space of unfettered flows, and her spatial hybrid of "place as meeting place" (p. 41). The first two approaches underpin modernization and globalization theses respectively, and they effectively hide spatial differences within their Whig histories. Massey's own spatialization of social analysis combines complex, layered histories with open, connected spaces as an alternative to simplistic "non-spatialised globalization" (p. 33). This is the complexity of globalization, which the editors claim their book is "a testament to" (p. 25). Unfortunately, the authors in the other three parts never refer to Massey's chapter, and only one makes even a brief mention of her earlier work on this subject (p. 137).

Part 2, entitled "Risk, Society and Governance," includes two chapters on "environmental issues" and one on how to do sociology. Ian Welsh describes an emergent "Green regime of accumulation" in which "northern knowledge" is used to legitimate "northern advantage" in a world market where environmental risks are minimized. In a fine complementary argument, Marie A. Mater provides a review of the "southern critique" (largely Indian) of the dominant discourse on global environmental change, which serves northern corporate needs. In a useful link, both of these chapters provide a critical reading of the 1992 Rio conference. In contrast, Roger Sibeon's chapter on governance uses EU decision making as the policy focus to illustrate his "anti-reductionist sociology" (p. 83; it is also antiessentialist, antireification, and antiteleological) for studying "postnational governance." As the latter is the management of networks under conditions of contemporary globalization (p. 87), there is an obvious potential to link with Massey's ideas via reference to international financial flows, but this is missed (p. 91).

Part 3, entitled "Synergetic/Discrepant Difference," comes closest to providing a sustained analysis of the globalization thesis. It comprises three chapters, each providing a divergent voice from traditional views of globalization. Ngai-Ling Sum investigates "othering" in post-Cold War

international economic relations through three "new orientalisms"; Barnor Hesse critiques the "repressive discourse of western globalization" (p. 125) for its historical forgetfulness with respect to imperialism and promotes, in its place, a "reflexive globalization" illustrated through the Black diaspora (p. 132); and finally, Thomas A. Acton focuses upon the way Romani organizations have been able to use the language of globalization to legitimate nonterritorial and multicultural identity through papal recognition of them as a "transnational" minority, thus challenging the absolute sovereignties of nation-states (p. 150).

Part 5, entitled "Migration and Globalization," comprises three chapters with different concerns for the movement of people. Robert Miles takes a political economy approach, contrasting the degree of freedom of movement between labor and capital with a focus on the former in the specific restrictions imposed at airports. This "concrete-isation" of globalization is continued by Claire Wallace on movements in post-Cold War Central Europe acting as a buffer zone between a stable West and the new unstable East. Cultural and ethnic linkages are emphasized in this account. The most concrete chapter of all is left to last: Jennifer Platt and Phoebe Isard remain in Central Europe but go back to when it was "Eastern Europe" during the Cold War in a study of the influence that Hungarian refugees had on British intellectual life after 1956. For me, this final chapter summed up my misgivings about this book. Although it manages to have the word globalization in its title, the study could well have appeared in many another collection of papers with no reference at all to globalization.

While there are many relevant points in these 11 chapters, the book never makes the transition from an interesting aggregation of conference papers to an integrated contribution for critiquing globalization: there is a lot of raw material, but the editors leave too much work to the reader. I think the time has come for this type of publication to be transferred to the Web, which is particularly suited to presenting the immediacy of ideas at conferences (lost in the three-year gap between conference and publication in this case), possibly including discussions. Such conference Web sites would not prevent the papers from appearing in a hard format as revised journal papers or research monographs, on which libraries would be better spending their money.

Marketing the American Creed Abroad: Diasporas in the U.S. and Their Homelands. By Yossi Shain. New York: Cambridge University Press, 1999. Pp. 311. \$54.95 (cloth); \$17.95 (paper).

Sarah V. Wayland
Brock University

Challenging the view that U.S. foreign policy has been damaged by ethnic influences, Yossi Shain argues that ethnic groups in the United States

wanting to influence American foreign policy must first demonstrate how the policies they advocate toward their homelands will disseminate American values and ideals abroad. A secondary argument is that ethnic mobilization around foreign policy benefits American civic culture by reinforcing the values of democracy and pluralism among ethnic groups in the United States. This compelling argument is certainly timely, and the book addresses many of the recent debates on the politics of multiculturalism and transnationalism in the United States, yet—as Shain himself is careful to point out—the debates over ethnic influence in American politics are part and parcel of American history as well.

The book is the product of extensive research from secondary sources, as well as limited interviews with contemporary ethnic activists. The first two chapters provide a survey of ethnic politics in American life, primarily in the 20th century, including the role that various diaspora communities such as Irish, Slovaks, Greeks, Serbs, Haitians, and Cubans have played in advocating self-determination and regime change in their homelands. Three case study chapters follow, each of which is chosen to illustrate various components of diaspora politics. In chapter 3, for example, Shain's assertion that diasporas have to redefine their domestic and foreign policy strategies in light of global changes relies on the case of Arab-Americans in the 1990s and their responses to the end of the Cold War, American intervention in the Persian Gulf, the rise of militant Islamicist movements, and other issues and events. The subsequent chapter considers how events abroad affect ethnic relations within the United States by examining the relations of African-Americans and American Jews. Along with other issues, African-American support for Palestinians in Israel and the close ties between Israel and the apartheid state of South Africa placed severe strains on ethnic relations between blacks and Jews in the United States, relations which subsequently have improved. Another chapter uses the case of Mexican-Americans to illustrate the positive as well as conflictual relations that emerge between a home country (Mexico) and its diaspora. In the concluding chapter, the author returns to the question of ethnic influence on American foreign policy, reasserting his view that this influence keeps the United States true to its foreign policy ideals of promoting democracy, pluralism, and human rights, as well as directly addressing the critics of this thesis.

The author, himself an Israeli who has lived extensively in the United States, conveys real enthusiasm for the subject matter and peppers his analysis with knowledgeable references to the prototypical diaspora—the Jewish diaspora—and relations between Israel and Jews overseas. He successfully highlights the interaction of domestic and foreign issues for ethnic Americans, including the dilemma of divided loyalties between America and the homeland (particularly acute when a nondemocratic regime is in power in the homeland), as well as other issues and identities that divide diasporas internally. Not every case fits neatly into the author's thesis that diasporas provide a moderating influence on homeland divisions, for example, those of Serbs and Croats in the United States (pp.

64–66), but even those groups are careful in articulating their interests to American decision makers.

Shain's previous writings, namely his *Frontier of Loyalty: Political Exiles in the Age of the Nation-State* (Wesleyan, 1989), influence the current work. Though he has ostensibly extended his focus from exiles to diasporas, defined as "a people with common ethnic-national-religious origin who reside outside a claimed or an independent home country" (p. 8), *Marketing the American Creed Abroad* nonetheless primarily considers the role of ethnic community leaders and politicians as opposed to "average" members of a diaspora community. This is to be expected since the study focuses on the political dimensions of transnationalism. However, it was disappointing that Shain did not rely on the transnationalism literature per se, despite the fact that a growing number of nonpolitical scientists write about the political dimensions of transnationalism. Shain's analysis could also have been strengthened by drawing from social movements literature—the research on framing and political opportunity structures in particular.

Other gaps in the book include Shain's omission of basic demographic data. For example, how many Arab-Americans are there? Where do they live? The author mentions that a majority of them are Christian (p. 124) but does not go into any detail about this interesting fact. Last, Shain claims that diasporas in the United States are increasingly influential (pp. 1, 51, and others), but his use of historical examples alongside contemporary ones would seem to contradict this assertion. Further exploration of the hypothesis that diaspora politics are being fueled by the forces of globalization or by post-Cold War policy confusion would have made this already noteworthy book an even more important contribution to the rapidly growing literature on diaspora politics.

Ambiguities of Domination: Politics, Rhetoric, and Symbols in Contemporary Syria. By Lisa Wedeen. Chicago: University of Chicago Press, 1999. Pp. 256. \$45.00 (cloth); \$17.00 (paper).

George Steinmetz
University of Michigan

In this fascinating book, Lisa Wedeen analyzes the effects of the personality cult of President Hafiz al-Asad on everyday political life in Syria. *Ambiguities of Domination* begins with the puzzle of the cult's very existence: How does the state benefit from a cult whose contents seem preposterous to most Syrians? In answering this question, Lisa Wedeen provides a historical account of the cult's emergence; she interprets its imagery and the critical and artistic responses to it; and she offers an exciting theoretical discussion of theories of political subjectivity under authoritarian rule.

The first section reconstructs the development of the cult's vocabulary

and grammar since the mid-1970s. Family and gender metaphors, capable of symbolizing both obedience and national belongingness, are preponderant. Asad is depicted as father, Syrian citizens as children, the Arab *umma* as woman. The president and his family are portrayed as sacred and immortal. Yet the cult's iconography is also ambiguous. Symbols often seem to contradict one another, or to disappear and reappear over time. Asad is represented simultaneously as the nation's premier scholar, teacher, and pharmacist; he is both a "knight of war" and a "man of peace." While the images of Asad often stress his "manliness," Wedeen notes that in actual public appearances he is "not particularly energetic," appearing stooped, hoarse, and sluggish.

Rather than seeing these inconsistencies as undermining the regime's effectiveness, however, Wedeen contends that the Asad regime actually *increases* its power by demanding that "citizens provide external evidence of their allegiance to a cult whose rituals of adulation are manifestly unbelievable" (p. 68). Compliance is not based on Weberian charisma, nor is it based on Gramscian hegemony, defined here as ruling ideas that set a limit on "what will appear as rational, reasonable, credible, indeed sayable or thinkable" (p. 11). Evidence from various sources and places suggests that many Syrians hold an irreverent and dismissive attitude to the cult. At the same time, however, they are "fluent" in the cult's iconography and language. The cult's political effectiveness derives from mandatory, ritualized practices in which citizens are themselves "accomplices." This isolates them from one another and clutters public space with "monotonous slogans and empty gestures." "People are not required to believe the cult's fictions, and they do not," Wedeen writes, "but they are required to act *as if* they did" (p. 30). She draws on Vaclav Havel's analysis of a similar "nonhegemonic" situation, state socialism, in which citizens mutually enforce one another's obedience. As Havel notes, people in such conditions tend to identify with their involvement as "something natural" and come to see noninvolvement as an "abnormality." Hence there is a kind of "naturalization," but it is naturalization at the level of "material practices" and not, *pace* Gramscian hegemony theorists, naturalization of the regime's substantive ideas.

The cult also invites transgressions. Wedeen demonstrates this in a long chapter analyzing the more or less oblique forms of political criticism in Syrian films, plays, televised comedies, jokes, and political cartoons. She argues against seeing such criticism as either a functional safety valve or as full-scale resistance. By operating within the grammar of the official political culture, such criticism actually *underscores* the cult's power. On the other hand, critical public culture *counteracts* some of the cult's atomizing effects and provides the foundations of "what might ultimately eventuate in organized opposition" (p. 131).

Wedeen turns to several disparate theoretical approaches to make sense of the subjective condition of this state, one in which people are neither "totally alienated" from the relations of domination nor "totally inscribed" within them. Referring to Althusser and Foucault, she suggests

that by compelling people to mouth ridiculous slogans and perform absurd material rituals, the cult "interpellates" them as disciplined and obedient subjects. Later she refers to writers like Elaine Scarry, who suggest that certain coercive state practices force people to surrender their very sense of self. It is unclear how this more deconstructionist diagnosis—the suggestion that Syrian politics actually "‘de-constitute’" individual selves—articulates with the Althusserian image of power as involved precisely in the *constitution* of the subject. In a third theoretical move, Wedeen turns to a rational choice model of "tipping" or "cascades," writing that "as more people assess that others will refuse to comply, political life can 'tip' from compliance to protest" (p. 152). This reference to a cognitivist, transparent model of subjectivity stands in uneasy tension with the earlier theories. If we take seriously the idea that Syrians are "interpellated" as complicit, might one not expect the resulting forms of subjectivity to color their assessment of the desirability of protest and change? And if we accept the argument about the disruption of the sense of self, it is difficult to imagine how one could engage in the rational sorts of decision making required by the tipping model without first reconstructing the self. The author provides a hint of such a reconstitutive psychological process in her discussion of the deatomizing effects of critical public culture. I would have liked her to deal more explicitly with the relations between these different theoretical approaches or to specify their differing points of application.

Ambiguities of Domination is an invaluable contribution to the increasingly subtle literature on everyday forms of power in authoritarian political systems. It exemplifies the rising interest in cultural analysis in the social sciences and should be read by all students of culture and politics.

Gendering Politics: Women in Israel. By Hanna Herzog. Ann Arbor: University of Michigan Press, 1999. Pp. xvi+292. \$54.50.

Emily Stoper
California State University, Hayward

In this study of women members of local councils in Israel, Herzog manages to combine a penetrating, yet jargon-free deconstructionist analysis with an empirical approach that includes questionnaires mailed to women elected to local councils between 1950 and 1983, in-depth interviews, use of "every bit of published information" on the subject, and secondary analysis of existing data. As of 10 years ago, only 8.5% of local officeholders in Israel were women, compared with 38.4% in Sweden (the highest in the world), 19.0% in Britain, 15.0% in Uganda, and 14.0% in the United States.

Herzog paints a picture of Israeli society that offers little hope for future change in this low level of representation. The historical and theoretical background presented in the book are entirely based on Western Eu-

rope and the United States, presumably because Israeli women have not written their own feminist history and theory. There is virtually no feminist movement in Israel, and feminism is almost universally regarded in a negative light.

Herzog deftly deconstructs the Israeli gender role system, showing how it relegates women to a socially constructed separate sphere, defines politics in a totally masculinized way, and forces women who enter politics to define their roles as nonpolitical and as an extension of the feminine sphere of home, family, and voluntary organizations. Even women's paid work, because it is performed in a highly gender-segregated workforce, is constricted as part of the feminine sphere, so that women and men can maintain the belief that everything women do is feminine. Women's volunteer work, although mostly done in professional and bureaucratically structured organizations, is not seen as helping qualify them for either paid work or political office. Only volunteer work that serves others, not political organizing work, is validated for women by the gender system.

The crippling effect on women's political success of all these unquestioned and well-entrenched thought patterns is predictable. Most of the Israeli women council members serve only one term and then decline to run again, often because they have little success in achieving the community service goals that they brought to the council and which, given the gender system, are the only legitimate roles they could have brought. Only six women have ever filled the paid position of council head.

What makes Israel so patriarchal? Herzog offers many reasons. First, the long-term conflict with the Arab nations tends to make the military the most prestigious institution in society and tends to reinforce a macho ethos. Male officers can often leapfrog into high-ranking political roles, but even high-ranking women veterans cannot translate their military experience into political success. The sense of being beleaguered also buttresses traditional family patterns, in part by putting social pressure on women to reproduce their ethnic group (the average Jewish woman in Israel has three children; the average Arab Israeli woman has even more). The importance of religious parties in Israeli politics also works against women; most Orthodox Jewish parties take a principled stand against women's participation in politics, as do most Muslim leaders. Only four Arab women have ever been elected to local councils in Israel; all of them were from Christian backgrounds, and three ran as Communists.

Israel has a parliamentary system with strong parties and proportional representation, carried out through multimember districts and party lists (ranked slates of candidates). The Labor Party guarantees women 20% of the places on each list (though not necessarily in a slot that has a realistic chance to win). Other women have had some success in running on independent lists with no ties to the national parties. Wilma Rule and others have posited that this type of system gives women an advantage, but this hypothesis is not borne out by the Israeli experience. The strong-party system creates an atmosphere of infighting and back-scratching (ex-

change of favors), which Israeli women find distasteful and at which they are at a great disadvantage. Herzog shows how even the Labor Party guarantee of 20% of the places is so grounded in the ideology of women's special roles that it may hold women back as much as it advances them.

This insightful study is a welcome addition to the large body of material on women's political and social roles in comparative perspective. Ironically, one gets the impression that the audience that will read and understand this book will be almost entirely outside Israel.

A Generation Divided: The New Left, the New Right, and the 1960s. By Rebecca E. Klatch. Berkeley and Los Angeles: University of California Press, 1999. Pp. xiv+386. \$55.00 (cloth); \$22.95 (paper).

David S. Meyer
University of California, Irvine

Months before President John Kennedy announced that a "torch had been passed to a new generation," members of the next generation were already trying to seize that torch and run with it—in opposite directions. In July of 1960, conservative Republican students, supported by pundit William Buckley, founded Young Americans for Freedom (YAF) with the intent of organizing and mobilizing young people in the fight against communism and for individual economic and political freedom. One month earlier, sponsored by the League for Industrial Democracy and inspired by activists Michael Harrington and James Farmer, college campus leaders founded the Students for a Democratic Society (SDS) with the intent of promoting individual liberties, political participation, and social justice. These groups provoked some of the tumult of the decade that followed and responded to virtually all of it. Young people in and around each organization forged individual political identities as activists through a wide range of collective action. Rebecca Klatch's primary interest is in how individuals develop such identities and how activists on both sides made sense of their lives and the world in the 1960s and the following decades.

Klatch has organized her book to highlight the voices of the activists. The bulk of her analysis comes from interviews of 74 activists, roughly equally split between the two organizations and between men and women. Less a representative sample than a collection of interesting, self-conscious, and articulate informants, Klatch uses the "life histories" of her subjects to examine the processes of individuals coming to terms with themselves as activists over most of a lifetime.

At once, Klatch has made a contribution by putting YAF and the conservative movement into a history of the 1960s, which is too often treated as a time of left-only activism. Further, by comparing the stories of the young conservative activists with those of their counterparts on the student left, Klatch offers a broader perspective on both the roots of political

activism and the effects of activism on a life. Finally, Klatch's treatment of these ends of the political spectrum shows clearly how the same issues and events passed through and influenced a generation.

This most welcome book raises critical issues about opposing movements in American politics. The war in Vietnam, for example, proved a divisive watershed for both organizations, although in very different ways. SDS was inundated with new members who opposed the war, but frequently did not share the developed commitment to "participatory democracy" or egalitarianism that united the early members. The explosive growth of SDS allowed some members to become progressively more insulated from mainstream politics and culture. SDS split into competing factions, divided on both political analysis and tactics, and fell apart at the end of the decade, with the war still in process and Richard Nixon in the White House.

YAF remained a small and relatively marginal organization on the right and ultimately managed the ongoing tension between libertarians and traditionalists (Klatch's terms), who were divided on the war generally and the draft in particular, by forcing out the libertarians. These factions would be reunited in the late 1970s and the 1980s around Ronald Reagan's presidency. Similarly, the counterculture swept through both organizations divisively, with traditionalists on the right and Maoists on the left reacting against the spirit of sex, drugs, and rock and roll.

The similarities between the activists are striking and also strikingly different from popular understandings of student protesters. Almost all of these young rebels extended, rather than reacted against, the politics of their parents and families. Young activists on both sides of the spectrum recalled growing up in families where politics and values were discussed at the dinner table. Although the intensity of commitment and choice of tactics created rifts within the families of some SDS members, in all cases, activism was seen as an important part of a reasonable adult life. After the 1960s, individuals continued their political activism on the left or right in the 1970s and beyond, as individuals sought meaningful and viable politics in changing political environments.

The differences also provide insights into the politics of dissent. It is not surprising that the conservatives were much more attracted to, and successful within, conventional electoral politics, while many of the SDS alumni struggled to find effective political niches in other movements: particularly feminism, community politics, social services, and academia. Also not surprising, those who had seen themselves as socialists and revolutionaries in the 1960s expressed more discomfort with acquiring the material trappings of adult middle-class life in the United States than did the advocates of traditionalism and free markets.

By extending the political range of dissident movements in the 1960s, as well as the varieties of political activism over a lifetime, Rebecca Klatch has effectively challenged scholars to look for a broader range of social movements and to search for the connections between the most dramatic episodes of protest.

Race, Money, and the American Welfare State. By Michael K. Brown. Ithaca, N.Y.: Cornell University Press, 1999. Pp. xxiv+381. \$55.00 (cloth); \$22.50 (paper).

Dalton Conley
Yale University

The premise of Michael K. Brown's well-researched book is that the politics of race interacted with the politics of fiscal restraint to engender the two-tiered, punitive welfare state that now distinguishes America from most other advanced, democratic capitalist countries. The salience of race to the development of the modern American welfare state is not a new theme. Jill Quadagno's tour de force, *The Color of Welfare* (Oxford University Press, 1994) has already covered this ground, as has some work by Theda Skocpol. The value added here is the addition of the word *money* to the title, that is, the analysis of antitax, antispending rhetoric as it fits into this puzzle. The creed of fiscal constraint, according to Brown, has dogged American social policy since the 1930s, and it has been blacks who suffer the consequences, one way or another, of such monetary "limitations."

During the 1930s, battles over the shape of American social policy pitted northern liberal Democrats against probusiness Republicans. The Republicans stood firmly against the New Deal, and in order to gain majority support, northern Democrats were more or less at the mercy of their southern brethren. This led to a disproportionately high commitment of federal infrastructure funding to the southern states. More important, however, it led to programs of temporary work relief (as opposed to guaranteed employment) and to public pensions and social insurance that were not universal in scope. The result of these choices was the prevention of African-Americans from fully benefiting from the initial promise of the New Deal. In this process, Brown sees fiscal restraints as more determinant than other institutional constraints (such as the potential unconstitutionality of universal employment policies, e.g.).

After World War II, it was again fiscal politics that resulted in the ultimate demise of Truman's Fair Deal plans for universal health care and other social-democratic, race equalizing benefits. Since Truman was battling inflation, he needed to complement his high spending policy with a high tax policy in order to constrain the money supply. However, when the 1949 recession hit, the old cries for reduced expenditures prevailed in Congress and the high spending part of his policy fell by the wayside. Likewise, during the 1960s, Brown argues that the Great Society was limited by the 1964 Kennedy-Johnson tax cut that was made to please business interests. Budgetary limitations—in combination with Vietnam—dogged Johnson for the rest of his presidency. In one instance, he even traded a \$3.1 billion decrease in the budgets of NASA, the Department of Agriculture, the Atomic Energy Commission, and Food for Peace in order to gain a \$3.3 billion increase in spending for Great Society programs. Such are the choices that the rhetoric of balanced budgets force upon us.

Ironically, it was during the heady 1960s that the underlying ideals of the New Deal were ultimately abandoned, according to Brown. For it was during the Johnson administration that New Deal liberals began arguing for means testing of government programs. Again, this was the result of the fact that, given overall budgetary constraints, liberals thought the best way to achieve their redistributive ends was to target their limited resources to the most needy. What they did not count on was the fact that the end of the promise of universalism was the death knell for large-scale social policy since "welfare" lost its potential to be conceived as a social right of citizenship. It also paved the way for an association of race and welfare in the public consciousness that was unprecedented.

The lens of fiscal politics that Brown provides is an elucidating one. However, there are some weaknesses to the text worth mentioning. For instance, given that the premise of the book is that the need to toe the budgetary line (or at least the appeal to such a need) was the central agent in the racial stratification and truncation of the American welfare state, the book's thinness regarding welfare policy during the Reagan years is a glaring weakness. It was during the 1980s that Reagan (and his budget director Stockman) cut a tacit devil's pact with Democratic congresses to simultaneously cut taxes and increase military spending without substantially reducing social spending. Despite such fiscal unrestraint and ballooning deficits, the racial stratification of welfare never seemed greater. How does this period in American social policy fit with the Brown thesis? Second, given that this is a book about race and the welfare state, and about the limitations placed on the power of government by the rhetoric of fiscal restraint, it seems surprising how cursorily Brown addresses the guns versus butter debate.

Despite such shortcomings, this is an important book for students of race and social policy. Perhaps Brown's greatest contribution will be in making sociologists think about budgets and fiscal rhetoric as a social institution and political structure—one just as important as subcommittee rules, a bicameral legislature, or the tripartite system of checks and balances in determining the possibilities and limitation of American social policy.

Being Black, Living in the Red: Race, Wealth, and Social Policy in America. By Dalton Conley. Berkeley and Los Angeles: University of California Press, 1999. Pp. 217.

David A. Reingold
Indiana University

Is it race or class that determines an individual's life chances? This is the primary question that Dalton Conley attempts to answer in *Being Black, Living in the Red*. Using a unique longitudinal data set known

as the Panel Study of Income Dynamics (PSID), Conley attempts to untangle the intergenerational effects of class and race, controlling for other basic socioeconomic characteristics, and estimates their relative impact on net assets, educational attainment, wages, premarital childbearing, and welfare receipt. His general hypothesis is that "it is not race per se that matters directly; instead, what matters are the wealth levels and class positions that are associated with race in America" (p. 7). His results indicate that the intergenerational transmission of poverty and affluence are more strongly influenced by parental assets, or class, than by race. The book concludes with a discussion of these findings and their implications for antipoverty programs, including race-based affirmative action policies.

While the book's central research question has been the subject of a copious and distinguished social science literature, Conley's analysis does advance the race versus class debate in several ways. First, the PSID provides longitudinal data on a random sample of American families. This allows for the tracking of individual social mobility over time with a representative sample that can be generalized to a much larger population. Second, the PSID provides information on the socioeconomic conditions of a respondent's household over the life course without having to rely on data that is based on the respondent's retrospective memory. This allows for more accurate measures of the socioeconomic conditions of households in which respondents were reared, as well as their current socioeconomic position as adults. More important, the structure of these data limit the problem of simultaneity. Third, the PSID collects information on household assets, including the value of real estate, bank accounts, stocks, and other forms of capital. As the author notes, this is an improvement over previous research, which has largely equated class with labor market position, including occupational status and household income.

Overall, Conley's findings suggest that "if we statistically compare blacks and whites who are similar in terms of their individual characteristics (age, gender, number of siblings, and, in some analyses, education and income levels), their family backgrounds (parents' age, whether they grew up in a female-headed household or one that used welfare), and their class origins (parents' education level and occupational prestige, as well as their family's permanent income, net worth, and the types of assets), . . . blacks are just as likely as whites to have completed college, . . . the black-white wealth gap among young adults disappears . . . , and the racial differences in the chance of using welfare among this age group also vanish" (pp. 133-34). While race, independent of class, still seems to play a role in the number of hours worked and out-of-wedlock births, "black workers who are employed full time enjoy a wage advantage over white workers; additionally, net of family background, the latest cohorts of African Americans are more likely to graduate from high school than their white counterparts" (p. 134). Therefore, current racial disparities in life chances are not so much the result of race, but a result of extremely

low levels of asset accumulation among blacks—in “1990 black Americans owned only a meager 1 percent of total wealth” in the United States (p. 25).

While these findings are compelling, the extent to which these results can be generalized to the entire U.S. population requires further comment. As the author notes, the PSID is conducted every year, starting in 1968, and follows 5,000 families selected at random, including new families that evolve out of the original sample. Since Conley is interested in measuring the impact of childhood family wealth on later life outcomes, he correctly restricts the sample to those children born since 1962. This leaves a working sample of 1,285 respondents; 11.1%, or approximately 143, are black. Although the remaining sample size is large enough to produce reliable estimators, it raises concerns that the findings suffer from a lack of external validity. That is, it is unclear whether these 143 cases are, in fact, representative of all African-Americans. Moreover, it remains unknown whether the number of black cases in the multivariate analysis is even smaller due to missing data since these figures are not reported.

The other curious aspect to Conley's book is the disconnect between his findings and the final chapter on public policy implications. Although this chapter explores past and current initiatives designed to increase asset accumulation among African-Americans, it virtually ignores the research findings' implications for race-based affirmative action policy. By sidestepping this politically sensitive issue, the author leaves it up to the reader to interpret his results for an important area of public policy. For social scientists interested in public policy, this can be a very dangerous strategy. After all, the obvious interpretation of this book is that race-based affirmative action should be abolished in favor of class-based affirmative action, where class is measured in terms of assets.

Overall, *Being Black, Living in the Red* is an important contribution to our overall understanding of social stratification in modern society. It provides an excellent introduction to the subject for students and is an important piece of scholarship for all social scientists interested in inequality.

Something Within: Religion in African-American Political Activism. By Fredrick C. Harris. New York: Oxford University Press, 1999. Pp. xii+227. \$35.00.

Mary Pattillo-McCoy
Northwestern University

Despite the fact that Frederick Harris is a political scientist, *Something Within* is a profoundly sociological book. The author draws on both the classical theory of Karl Marx and Émile Durkheim, as well as the contemporary work on (black) religion and political activism by sociologists

(and a political scientist) Cheryl Townsend Gilkes, David Snow, Gary Marx, Patricia Hill Collins, Aldon Morris, and Sidney Verba. Harris combines an impressive variety of data—surveys, polls, historical sources, autobiographies, and rich ethnographic descriptions of Chicago's black political-religious scene—to examine the psychological, institutional, and cultural "resources" that inhere in black churches and are enacted by black Christians. His empirical investigation of the workings of these resources brings him to the conclusion that "far from subverting black political activism, Afro-Christianity stimulates black political activism" (p. 177).

First, Harris shows that black religion supplies the fortitude and feelings of self-efficacy necessary to organize and act. This chapter on psychological resources is entitled "Blessed Assurance," which is taken from a hymn, and the chapter opens with a quote from Ethel Gray, a Civil Rights organizer in Mississippi, who recalled, "We stood up. Me and God stood up" (p. 69). Harris's use of such elegant fragments from religious songs, words of wisdom by black activists, and biblical verses makes this book especially inspiring and provides powerful illustrations of his own theoretical contributions (e.g., his explication of "religiously inspired political efficacy.")

The next resource that Harris studies is the institutional strength of the church in a chapter called the "Rock in a Weary Land." Most illuminating is his argument that during the post-Reconstruction "nadir" of race relations, black churches turned inward where members honed the political and organizational skills that would later be activated in the more sympathetic opportunity structures of the 1950s and 1960s. Using historical sources and contemporary ethnographic data, Harris emphasizes the networks of black churches and the political information disseminated through them. Then, using survey data, he finds that blacks are more likely than whites to approve of religious leaders being politically involved, report more political campaigning activity at their churches, and are more likely to discuss politics among themselves in a church setting.

Finally, Harris analyzes Afro-Christianity's oppositional culture, which provides the frames and language for political action. From Nat Turner's slave rebellion to contemporary Chicago politicians who use biblical verses and imagery to liken themselves to Jesus, the cultural content of black religion has motivated political action. There are many ethnographic treasures in this chapter, such as the following prayer for an African-American mayoral candidate in Chicago: "Father God we ask that Thou will anoint for us the head of our brother Danny K. Davis that Thou will give him the wisdom of a Solomon and the strength of Samson, Lord" (p. 146).

This last resource is a perfect example of the themes of paradox, duality, and simultaneity that run throughout *Something Within*. While the black church exudes an oppositional culture, it simultaneously acts to hold church members within the bounds of legitimate forms of action. Harris argues, "The culture and institutions of dominated groups can

nurture political beliefs and strategies that at the same time both support and oppose the existing civic order" (p. 67). Similar paradoxes exist with respect to the role of women. The black church stridently fights racial domination but itself promotes (implicitly and explicitly) gender domination. The church has provided an important forum for black women to learn and practice organizational and leadership skills in their own auxiliary organizations, but these separate and sometimes subordinate locations can negatively affect the understandings of the role of black women in secular politics. Indeed, Harris finds that "black women's support for feminist principles softens among black women who are the most active in their churches" (p. 168). A final paradox is that while the black church has fought for inclusion, its theocratic tendencies have also "undermined democratic processes in African-American politics and society" (p. 179). Thus, in the types and forms of political engagement, and regarding gender, the components of Afro-Christianity are sometimes paradoxical and always complex and multifaceted.

The one critique of this well written and well argued book is perhaps just a rhetorical quibble. Harris introduces the book by stating: "The story I am about to tell is highly controversial" (p. 4). While there exists no consensus about the role of the black church in politics, and there remain those who insist that Afro-Christianity retards black activism and is unredeemably sexist, the fact that Harris could cite so much literature that buttresses his arguments indicates that few students of the black church think that it is simply a placating "opiate." And the camp of those who recognize its institutional, cultural, and psychological (to a lesser degree) resources is definitely growing. Nonetheless, Harris has made an important contribution to this field.

Faithful and Fearless: Moving Feminist Protest inside the Church and Military. By Mary Fainsod Katzenstein. Princeton, N.J.: Princeton University Press, 1998. Pp. xiv+270. \$17.95.

Mary Bernstein
Arizona State University

In lively and engaging prose, Mary Fainsod Katzenstein's *Faithful and Fearless* uncovers new realms of feminist protest inside two unabashedly nondemocratic, hierarchical, and patriarchal institutions—the American military and the Catholic Church. Katzenstein argues that access to the law, activists' accountability to their institutions, and exposure to radical ideologies determine the vastly different forms feminist protest takes within institutions. Katzenstein redefines protest, expands an understanding of contemporary feminism, and theorizes the relationship between law, forms of protest, and social change.

Katzenstein's carefully researched analysis moves beyond the often sterile definitional debates in social movement theory to interrogate the

meaning of protest in an empirically grounded way. According to Katzenstein, "I use the term *protest* despite the fact that the women whose activism I describe are far from lawless, rarely use civil disobedience, and never resort to violence. Less lawbreaking than norm-breaking, these feminists have challenged, discomfited, and provoked, unleashing a wholesale disturbance of long-settled assumptions, rules, and practices" (p. 7). She further distinguishes the "discursive politics" of feminists in the American Catholic Church from the "influence-seeking interest-group politics" of feminists in the military. Whereas the former is concerned with meaning making, "deploy[ing] language and symbols to convince others of new possibilities" (p. 19), the latter engages in moderate politics-as-usual.

Katzenstein first describes then analyzes the development of feminisms within the American military and the Catholic Church. Adopting liberal, interest-group strategies, military feminists demand equal pay for equal work, an expansion of military occupations available to women, and an end to sexual harassment. Framing claims in terms of enhancing military preparedness, they are willing to stay silent or make compromises in order to meet their goals. Military feminists have been slow to embrace issues of race discrimination and were mostly unwilling to form coalitions with groups seeking to end the bar on homosexuals in the military. In contrast, feminists in the American Catholic Church engage in discursive politics, creating new words and meanings "to describe their changing understanding of women in society and church" (p. 123). Emphasizing radical equality, feminists have moved beyond simply calling for women's ordination in classic equal opportunity fashion to "combine . . . a critique of societal inequalities, racism, patriarchy, imperialism, and militarism with a sharply etched critical analysis of institutional hierarchy and clericalism" (p. 133). Although not always successful, feminists within the American Catholic Church actively pursued a politics of antiracism, sought to end homophobia, and worked on behalf of the poor.

Katzenstein's most significant contribution comes in explaining the development of these different forms of feminism. According to Katzenstein, "Over the last decades the incorporation of rights language within the law and the broader institutional diffusion of equality of opportunity norms have led to the 'legalization' of protest inside institutions" (p. 35). Beyond simply providing a political opportunity, the law shapes the values, beliefs, and preferences of activists. Strategies and outcomes are therefore context specific, shaped by the broader political environment, the norms of the institutions in which activists are located, and recourse to the law. Because the law curtails claims about the intersectionality of oppressions, little incentive remains for military feminists to seek alliances or to make claims based on race or sexual orientation. The minimal use of discursive politics by military feminists and the embracing of equal opportunity norms also means that military feminists never "fully confront . . . the question of whether the maintenance of an efficient fighting force requires adherence to traditional masculinist models of mil-

itary prowess" (p. 81). In other words, the very responsiveness of the law to equal opportunity claims narrowed the agenda of military feminists, while the lack of access to the law pushed feminists in the church in a more radical direction.

Political success must be measured in a variety of ways. Far from opting out of politics, discursive strategies challenge traditional power arrangements. Although the temperate politics of military feminists resulted in the expansion of military occupations open to women, sexual harassment persists, women are still excluded from combat, and uniformed men's use of prostitutes continues unchecked. In contrast, women in the Catholic Church may be no closer to ordination today than they were 30 years ago, but feminist discursive politics has "impeded the 'mobilization of bias' that would have allowed the hierarchy to keep women's issues off the agenda for public debate" (p. 168) and "has clearly led to a swing in Catholic lay opinion toward feminist positions" (p. 169).

This impressive piece of scholarship should point researchers toward examining the context in which different political strategies arise and to critically evaluate the types of outcomes that result. Katzenstein's astute analysis of how the proximity of institutions to the state and the law shapes protest should provide a stimulus to future research. This study could be used in a variety of graduate and undergraduate courses, including courses such as social movements, political sociology, and law and society, and is a must for social movement theorists and feminist scholars alike.

Ordaining Women: Culture and Conflict in Religious Organizations. By Mark Chaves. Cambridge, Mass.: Harvard University Press, 1997. Pp. x+237. \$29.95.

R. Stephen Warner
University of Illinois at Chicago

About half of all Christian denominations in the United States now grant full formal leadership status to qualified women (that is to say, they "ordain women"), whereas half do not. Those that do so include the Congregationalists, Episcopalians, and Unitarian-Universalists, many of the African-American churches, the Northern Baptists, and the "mainline" branches of the Presbyterian, Methodist, and Lutheran churches. Those that do not include the Roman Catholic, Southern Baptist, and Missouri Synod Lutheran churches. The first group to take this step (the Congregationalists) did so in the middle of the 19th century, and the latest (including the Episcopalians) did so within the past quarter century. The rate of innovation—denominations newly adopting women's ordination—was rapid at two phases of the process, in the 1890s and 1970s, but slowed, even reversed, after 1980. Ordaining women became contentious at the end of the 20th century, which provoked Mark Chaves to find out

what the fuss was all about. Perhaps the easiest way to think about the resulting book is that it proves, through sustained argument, that its main title is seriously misleading. "Conflicts about women's ordination are about gender, not women" (p. 189)—so the author sums up his thesis—and thus the book's topic is not "ordaining women," a topic on which there is a literature, but instead deciding whether the status of ordained clergy can apply to women, whether or not there are any women candidates at hand.

Brilliantly drawing on the perspective of the new institutionalism, Chaves shows convincingly that the "pragmatic" issues of whether religious organizations need and use women's talents in responsible roles and whether women are available for and eager to occupy such roles are largely independent of, at best "loosely coupled" with, the issue of whether religious organizations are willing to endow women with formal authority. Formal authority turns out to be increasingly a "symbolic" issue, less about the women who actually do or would like to do religious work than about messages the respective denominations intend to communicate.

For most of this century, those denominations that have decided to make ordination available to women have meant through that offer to signal their endorsement of "modern" norms, particularly the application of "Enlightenment humanism" to the gender order, so Chaves claims. Thus the mainline Methodists and Presbyterians granted formal clergy rights to women already in the 1950s, well before being pressured to do so by second-wave feminism or by demands for recognition on the part of the women graduates of their seminaries, who began to appear in large numbers only in the 1970s. Conversely, those denominations that still refuse to ordain women, even those, like the Catholic church, that are increasingly dependent on women as *de facto* clergy, intend by their refusal to signal their resistance to modernity and all its works, most especially to the feminism that is so deeply a part of it.

In the longest chapter, "Inerrancy, Sacramentalism, and Women's Ordination," Chaves sets himself a complicated twofold task. He identifies the two religious audiences—networks of Biblically inerrantist and liturgically sacramentalist denominations—that constitute the reference group with respect to which those who refuse ordination intend to make this symbolic gesture. He then refutes the view, widespread among both partisans to and observers of the ordination conflict, that either the Bible itself or the nature of the sacraments logically prohibit women from being pastors or priests. In my opinion, Chaves's refutation is more convincing with respect to the inerrantist denominations, some of which have in fact reversed course on the matter while their Bible stays the same; thus, refusing women's ordination is a new litmus test of conservative Protestantism. It is less clear which denominations constitute the interested audience that the Vatican needs to placate. Who are the Joneses that Pope John Paul is trying to keep up with? It is also unlikely on the pro-ordination side that any urge to endorse the Enlightenment project explains

the much earlier adoption of women's ordination by such holiness and Pentecostal groups as the Salvation Army (1870), the Church of the Nazarene (1908), and the Assemblies of God (1935) than that of the northern Presbyterians (1956) and the Episcopalians (1976; these dates are from table 2.1, pp. 16–17). Framing women's ordination as a referendum on modernity is a quite recent contribution of the women's movement to the construction of the two-party system of white American Protestantism.

There is much else in the book to ponder, both appreciatively and critically. The main thrust, the insightful and clearly written explanation of the ordination controversy in terms of the new institutionalism, makes the book a major contribution to two fields, sociology of religion and of organizations, as well as to their growing articulation, amply justifying its selection by the Society for the Scientific Study of Religion for their 1999 Distinguished Book Award. It is a bonus that the book is meticulously edited and beautifully bound.

Congregations in Conflict: Cultural Models of Local Religious Life. By Penny Edgell Becker. New York: Cambridge University Press, 1999. Pp. xii+267. \$59.95 (cloth); \$18.95 (paper).

Mitchell L. Stevens
Hamilton College

Penny Edgell Becker began her study with a series of empirical questions. What factors predict conflict in local religious groups? What do dispute trajectories in congregations look like? Is there variation in the nature of conflict by a group's size or political structure? To get answers Becker spent a lot of time in the western Chicago suburbs of Oak Park, River Forest, and Forest Park, learning about conflict in congregations that vary by faith, cultural orientation, size, and political structure. Her findings, some tellingly counterintuitive, are important for scholars of U.S. religion and the culture wars and are good news for institutional sociologists eager to expand their empirical vision.

Becker learned primarily by talking to people. She conducted over 230 interviews with clergy and lay people about the disputes in their own churches, parishes, and synagogues. She also worked as a participant-observer, sitting in on weekly services and coffee hours. Coding the data yielded 65 conflict events in 23 congregations. The puzzle is that neither a congregation's size, nor its political structure, nor a liberal or conservative cultural orientation robustly predict the presence or kind of conflict in particular congregations.

The key is that congregations have very different conceptions of what they are up to, different ways of explaining (in the pregnant vernacular Becker uses to describe her central analytic) "who we are" and "how we do things here." Across the 23 congregations, Becker found four congrega-

tional models—characteristic ways congregations make sense of themselves: *houses of worship* imagine themselves to be primarily about providing worship events and religious education; *family* congregations are primarily interested in providing meaningful interpersonal ties; *communities* care a lot about open debate on congregational issues and about inclusive decision making; *leader* congregations place considerable importance on activism and prominence in the surrounding locality. Finally, a few of Becker's cases are mixed congregations, in which no single model is dominant.

Variation in how congregations imagine who they are and how they do things goes a long way in predicting the amount and kind of conflict they experience. For example, houses of worship have the least amount of conflict because their members have rather thin expectations about what the congregation is supposed to be up to; there is relatively little to disagree on. Community and leader congregations have greater amounts of conflict, but their patterns of dispute resolution are different. In community congregations "the objective is to come up with either consensus or compromise that incorporates everyone's valued objectives." But in leader congregations "the process is not the point. . . . The outcome is what matters" (p. 143). Most notably, congregations in which no single congregational model is dominant have the most conflict. In these groups, not only the explicit terms of the dispute (e.g., about whether to spend limited funds on building maintenance or social service programs), but also the framework for making sense of it is up for grabs. Like soured relationships, in which surface troubles reflect much deeper divides, mixed congregations lack the shared purpose that make any particular tension easier to resolve.

There is a lot here for scholars in several fields. To sociologists of religion who variously describe and decry the rise of individual expressive styles of religious adherence, Becker says that the trajectory of this change is not unilinear. Her mostly white, middle-class suburbanites do have a voluntaristic understanding of their faith, but they are committed to their congregations nevertheless and in multiple ways. To "culture wars" analysts who posit a divide between orthodox and progressive cultural orientations among America's religious, Becker retorts that organizational sensibility trumps cultural orientation as an explanation for within-group conflict. And for the new institutionalists in organizational sociology, Becker's work demonstrates how portable their notions of model, metaphor, and logic of action can be.

There is less here for undergraduates or for generalists looking for a good read. At times, Becker gets trapped in her discipline's elaborate prose. Phrases like "publicly institutionalized rhetorics of discourse" (p. 225) confuse rather than clarify ideas and would make the book a hard sell in most classrooms. But *Congregations in Conflict* was not written for them. Among academic professionals, those interested in contemporary religion, group conflict, and cutting-edge issues in culture and organization will find the book well worth a careful reading.

The author's most remarkable accomplishments have been to reveal empirical models of organization that are often invisible even to their adherents and to show that variation among models has formidable explanatory power. The next chapter in this inquiry, one rightfully beyond the scope of this book, is to figure out where the models came from. From where did Becker's suburbanites get their metaphors of churches as families, communities, or houses of worship? Where and when were these metaphors worked out in our religious past? Have earlier faithful imagined their congregations differently, and if so, why have some metaphors survived while others have withered? Such questions bring Becker and her readers to intriguing intersections of sociology, religious studies, and cultural history.

After Heaven: Spirituality in America since the 1950s. By Robert Wuthnow. Berkeley and Los Angeles: University of California Press, 1998. Pp. ix+277. \$29.95.

Peter Dobkin Hall
Yale University

This study of contemporary religiosity argues that a "profound change in our spiritual practices" has taken place during the past half century. This change is a distinctive shift involving a "subtle reordering" in "how Americans understand the sacred itself" and centering on the replacement of the "traditional spirituality of inhabiting sacred spaces" with a "new spirituality of seeking" (p. 3).

Traditional spirituality, according to Wuthnow, emphasizes *habitation*—"the notion that God occupies a definite place in the universe and creates a sacred space in which humans can dwell as well" (p. 3). In contrast, the "spirituality of seeking" emphasizes *negotiation*—in which individuals seek those fleeting but "sacred moments that reinforce their conviction that the divine exists." Within this framework, people "explore new spiritual vistas," negotiating "among complex and confusing meanings of spirituality" (p. 4).

Wuthnow ties changes in spirituality to the social, economic, cultural, and political transformations of American life in the postwar decades: geographical, occupational, and economic mobility; increasing levels of education; and suburbanization that transformed faith from "something that people inherit" to "something for which they strive"—and do so increasingly as isolated individuals rather than as members of extended families and stable communities (p. 8).

"For seeker-oriented spirituality," Wuthnow suggests, "the congregation is less aptly characterized as a safe haven" than as "a supplier of spiritual goods and services." Large congregations often offer wider product lines—self-help groups, youth activities, bowling leagues—in order "to lure customers to come and to attend often in search of different grati-

fications," while small congregations function like boutiques, "becoming very good at something special, such as a ministry to the homeless, music, or sponsorship of a peace coalition" (p. 15).

Neither alternative, in Wuthnow's view, is entirely satisfactory: habitation spirituality encourages unrealistic dependence on "communities that are inherently undependable" and tends to replace worship of God with veneration for buildings; a spirituality of seeking, on the other hand, "is invariably too fluid to provide individuals with the social support they need or to encourage the stability and dedication required to grow spiritually and to mature in character" (pp. 15–16). Wuthnow offers a third alternative—a "practice-oriented spirituality" that puts "responsibility squarely on individuals to spend time on a regular basis worshipping, communing with, listening to, and attempting to understand the ultimate source of sacredness in their lives." Inevitably social and embedded in religious institutions, these practices, he argues, "must also be performed individually if they are to be personally meaningful and enriching" (p. 16).

Wuthnow's second chapter, "In the House of the Lord," provides a fine overview of the theological, ecclesiological, and socioeconomic characteristics of the golden age of denominational mainline religion in the late 1940s and early 1950s. Emphasizing the close connections between family and congregation, the geographic centeredness of communities, the domestication of the male breadwinner, and high levels of civic engagement, he concludes that "the local enclaves in which people worshipped were sacred fortresses . . . sacred space in which spirituality and identity were forged together," in which Americans "saw their country fulfilling a divine mission in the post-war world" (p. 39).

Wuthnow acknowledges that the habitation-oriented spirituality of the 1950s lacked depth, discouraged self-reflection, and denied the realities of evil embodied in the Holocaust and in racial segregation. In introducing the first of the finely textured interview-based profiles that comprise the balance of the book, however, he affirms his admiration for white, middle-aged, middle-class Midwesterner "Ned Stewart's" efforts to give his son "the same kind of sheltered spiritual world" in which he himself had been raised. At the same time, Wuthnow suggests that many of the difficulties Stewart encountered in doing so were due as much to shifting patterns of family and community life as to "changes in established religion" that made spirituality of dwelling harder to maintain as small-town churches relocated to growing suburbs and transformed themselves into large and impersonal "megachurches" serving geographically disparate congregations (pp. 43–44).

The balance of the book's seven chapters explore the varieties of spiritual journeys undertaken by postwar Americans as they availed themselves of new spiritual freedoms, including social activism of the right and left, the embracing of "family values" and other forms of spirituality based on discipline and asceticism, new age and charismatic enthusiasms,

the therapeutic quasi-religiosity of self-help groups, and, finally, the "practice-oriented spirituality" that he views as the most acceptable alternative to the religious dead ends of irrecoverable spiritual dwelling places and an unfulfillable spiritual searching. This tour through the spiritual possibilities of our time brilliantly interweaves ethnography with social and historical analysis of the ways in which religious leaders, ideas, and institutions struggled to meet kaleidoscopically shifting spiritual preferences.

For readers who have shared in the postwar generations' spiritual odysseys, Wuthnow's impassioned concern and scholarly daring will be immensely appealing. Others, more accustomed to the conventions of analytical and descriptive social science on which Wuthnow drew so ably in *The Restructuring of American Religion* (Princeton University Press, 1986), may find this volume puzzling.

Codependent Forevermore: The Invention of Self in a Twelve Step Group. By Leslie Irvine. Chicago: University of Chicago Press, 1999. Pp. vii+210. \$22.00 (cloth); \$15.00 (paper).

Diane Bjorklund
Illinois State University

In the last two decades, the term "codependency" has gained wide currency, with advocates claiming that up to 96% of Americans suffer from this "condition." Yet, at the same time, there has been abundant criticism of the notion of codependency as ill-defined, narcissistic, and replete with a victim mentality. In *Codependent Forevermore*, however, Leslie Irvine offers a defense of how the discourse used in a codependency self-help organization enables group members to make sense of their lives. Irvine conducted a 17-month ethnographic study of Codependents Anonymous (CoDA)—"a fellowship of men and women whose common problem is an inability to maintain functional relationships" (p. 7). CoDA was founded in 1986 as a 12-step organization and now includes roughly 4,000 groups in the United States. Irvine did most of her participant-observation in one group in Long Island, although she visited 17 other groups.

A fundamental tenet of CoDA is that people who are too eager to please others will lose their sense of self and an appreciation of their own needs and preferences. The solution, therefore, is for members to learn to "get in touch with their true feelings" and discover the "real self." CoDA differs from Alcoholics Anonymous by assigning the blame for problems not on the self as the source of its own troubles, but on the institution of the family as a transmitter of societal norms and standards that repress individuals' real selves. CoDA does maintain, nevertheless, that "recovery" remains an individual project (albeit while seeking the

company of fellow codependents) instead of calling for changes at an institutional level.

Irvine's primary point is that CoDA helps its members by supplying a "ready-made way of thinking" about their lives (p. 85). With its vocabulary of phrases, such as "making anger your friend," "old tapes," and "garbage guilt," CoDA offers its members a template for a coherent, respectable, and exonerative account of their lives (particularly their failed relationships). Echoing the ideology of the organization, members talk about enduring "emotionally abusive" parents during childhood, learning to recognize their real selves, developing personal standards for moral behavior by consulting their feelings, and gaining self-esteem.

In the latter half of the book, Irvine illustrates four narrative strategies used by members to describe their lives. Among those she interviewed, some tell stories that reconcile their desire for future relationships with their newfound emphasis on personal autonomy and their worry that a relationship may mean living life again on someone else's terms. Other members adapt the catch-all concept of codependency to talk about recognizing and managing emotions that result from relationships (particularly anger, love, and guilt), to describe their fledgling expectations of egalitarian gender relations, or in a few cases, to relate stories of being hapless victims of parents who failed to give them the unconditional love that was their due.

Irvine primarily concentrates on CoDA, although she does attempt some speculation on codependency as a cultural phenomenon reflecting contemporary changes in American selfhood and social relationships, despite her observation that CoDA's appeal is primarily to white, middle-class baby boomers. The broader implications of the emergence of the rhetoric of codependency, however, are developed more thoroughly in John Steadman Rice's *A Disease of One's Own* (Transaction Publishers, 1996), a book on CoDA that overlaps with Irvine's interests. Irvine coins a Madison Avenue-like phrase, "the institution lite," to delineate how an organization like CoDA can become a relatively undemanding replacement for traditional institutional anchors for identity such as lasting marriages and secure jobs. From her perspective, CoDA supplies a shared system of beliefs for making sense of self and relationships in an era of high divorce rates and unstable relationships.

Irvine includes a methodological appendix in which she explains her misgivings with across-the-board requirements of informed consent during fieldwork. She makes a controversial claim that revealing herself as a researcher was impossible in a group with no gatekeepers to grant permission for a study and where an announcement to the group would have disrupted the meeting. Thus she contends that she was thwarted in gaining consent from the whole group (in which she participated and shared her personal problems) until she was well into her study.

Codependent Forevermore will be of interest to researchers who study "autobiographical occasions," the instances when individuals tell stories about their lives, or to those who are interested in 12-step programs. Soci-

ologists are also likely to find this group intriguing because of the way in which members grapple with the tension between self and society—asserting their autonomy while acknowledging the obligations of living with others and the need for companionship. Irvine was wise to recognize that the rhetoric of codependency merits serious analysis, for it is a popular contemporary critique (judging by book sales on the topic) of American family dynamics and societal norms.

The Most Beautiful Girl in the World: Beauty Pageants and National Identity. By Sarah Banet-Weiser. Berkeley and Los Angeles: University of California Press, 1999. Pp. xiv+277. \$48.00 (cloth); \$17.95 (paper).

Maxine Craig
University of Papua New Guinea

In the introduction to *The Most Beautiful Girl in the World* Banet-Weiser notes that fellow academics often characterized her research as “fun” (p. 4). “Fun” was their way of saying that beauty pageants were trivial events whose interpretation required no scholarly effort. Banet-Weiser proves them wrong in a book that illuminates the beauty pageant as one of the best places to view the tensions surrounding the construction of gender. Her work is based upon interviews and ethnographic fieldwork in local and regional pageants in California.

Framing her book as a critique of those critics who dismiss pageant contestants as cultural dupes, she asks why women participate in beauty contests. Though she challenges what may be a popular feminist reading of beauty contests, her answer incorporates the work of a number of feminist scholars including Susan Bordo, Patricia Hill Collins, Kimberlé Crenshaw, and Jackie Goldsby.

According to Banet-Weiser, the Miss America Pageant works vigorously to prove that it is not what it appears to be. Officials insist that the pageant is not exclusive and most definitely not about sex. The pageant constructs an ideal of femininity fraught with contradictions. Its producers negotiate these contradictions through the fragmentation of femininity. The swimsuit, interview, and talent competitions separate contradictory aspects of femininity into distinct venues of performance. Contestants are bodies in the swimsuit competition and opinionated citizens in their interviews. They are noncompetitive but work tirelessly to win, spontaneous but cultivated, inviting but nonsexual, and committed to a better world but noncontroversial. Banet-Weiser argues that this impossible set of attributes adds up to an ideal of white middle-class femininity.

In chapters on the first Jewish, first black, and first disabled winners, the reader can see how the Miss America contest employs a liberal narrative of individual achievement to publicly resolve tensions surrounding multiculturalism. The pageant profitably manages diversity by simul-

taneously selling and denying the existence of difference. Race is introduced as a new and spicy element, while black contestants work overtime to distance themselves from the "social meanings of blackness" (p. 137).

Banet-Weiser introduces the reader to contestants, coaches, and swimsuit saleswomen but has very little to say about the spectacle's audience. In an otherwise perceptive chapter on the first black Miss America, the analysis loses its edge when Banet-Weiser writes about "the public" and how it interpreted the foreshortened reign of Vanessa Williams (p. 145). Very likely black, white, male, female, gay, and straight audiences had different readings of her crowning and scandalous downfall. The author shows the great effort that goes into dispelling the image that the Miss America pageant is a mere "leg-man's spectacle" (p. 38). For whom is respectability performed and with what effect? While this issue is tangential to some parts of her argument, it becomes more germane when Banet-Weiser addresses the meaning of the pageant. A glance at some of the sponsors she mentions—Clairol, a swimsuit manufacturer, and Cheer detergent—suggests that the display that critics have decried as a meat market is watched by an audience of women.

Banet-Weiser returns to the question of agency in her conclusion. Pageants, she writes, provide "sites for the construction of female liberal subjects" (p. 207). Here lies the potential for agency. Earlier in the book, she wrote, "Part of what the pageant provides is a place where women can negotiate a sense of self in relation to their bodies" (p. 69). Her evidence suggests to this reader that the pageant offers not just a place, but a particular way and stringently limited terms with which to negotiate a sense of self. Banet-Weiser reports that most contestants say that the best thing about being in a pageant is gaining "the ability to speak" (p. 93). In the context of the pageant, learning to "speak" is learning to give pleasing answers through the practice of what Banet-Weiser calls "rehearsed spontaneity" (p. 90). Later she reports that "what made it all worth it" for one winner was having a little girl recognize her and call out "you're SO pretty" (p. 101). It seems that the singular pleasure of being a contestant is a feeling of empowerment that is derived from the successful enactment of a valorized femininity.

Banet-Weiser concludes with a challenging theoretical maneuver. Having used feminist theory to criticize beauty pageants, she suggests that the reader use the beauty pageant as a mirror to hold up to liberal feminism. Recent pageants have appropriated feminist rhetoric as contestants present themselves as independent, career-oriented women who have chosen to take the opportunities provided by beauty contests. Perhaps pageants have been able to incorporate feminist rhetoric so easily because of the way liberal feminists have imagined agency as individual achievement. It is a deft and provocative ending to a thoughtful and readable book. Scholars interested in the dynamics of gender and race in popular culture will find the book especially useful.

Teaching America about Sex: Marriage Guides and Sex Manuals from the Late Victorians to Dr. Ruth. By M. E. Melody and Linda M. Peterson. New York: New York University Press, 1999. Pp. xii+287. \$28.95.

Julia A. Ericksen
Temple University

Melody and Peterson trace America's changing sexual mores over the past 120 years using marriage guides and sex manuals as evidence. Their book is in the tradition of much postmodern research on sexuality. Heavily influenced by Foucault, Melody and Peterson see these texts as creating sexuality and indeed gender. They examine the works for their disciplining and normalizing tendencies, and their main thesis is that despite seemingly liberalizing trends throughout much of the 20th century, the manuals have continuously eroticized masculine dominance in sexual matters.

The book follows a historical format starting with late Victorian sex and health manuals. I found this the most entertaining section, with its accounts of physicians who viewed the sexual health of individuals and the economic health of the nation as intimately intertwined. The authors move quickly to a chapter on Margaret Sanger. They see birth control as fueling America's first "sexual revolution" in the 1920s as well as the more familiar "revolution" of the late 1960s and 1970s. Other writers rating a chapter to themselves have been similarly influential in shaping American sexual patterns: Theodore Van de Velde, David Reuben, Alex Comfort, and Dr. Ruth Westheimer.

In the later chapters, an interesting discussion of gay and lesbian sex manuals written in the manner of Comfort's *Joy of Sex*—a book that the authors persuasively argue idealizes masculine dominance—underscores the conflict between gay male and lesbian views of sex. Indeed, Melody and Peterson state that only lesbian sex manuals develop a vision of sexuality based on intimacy and the whole body rather than power and the phallus.

Each chapter contains an introduction to the texts to be examined, followed by a detailed description of each writer's work—in some chapters there are several writers under discussion—and ending with a brief summary. Unfortunately, this leads to repetitive descriptions often lacking in historical context. Melody and Peterson rely too heavily on the excellent but already dated work of John D'Emilio and Estelle Freedman (*Intimate Matters: A History of Sexuality in America* [Harper and Row, 1988]) for their understanding of sexual history. Furthermore, their long descriptions of the selected books made me think that I would have preferred to read sections of the originals with perhaps a brief commentary. The authors say little about their criteria for inclusion beyond noting that they did not use sampling and instead used work that had been cited or

had gone through several printings. Each reader will find favorites left out—two in my case are Winfield Scott Hall and Judge Ben Wainwright.

The book is well written and readable, and I imagine it would work well in an undergraduate class on sexuality. One irritating aspect of the writing is the frequent foreshadowing of authors whose work is to be featured in a later chapter. A good editor could have caught this. For example, Dr. Ruth appears at least 18 times in chapters before the one discussing her work. Most of these occasions refer to the conservatism we are told we will find in her work, when we get to its discussion. Sometimes they even cite the work of writers from a later era whose work will not be featured. These kinds of comments are often made in a manner that is gratuitous to the point at issue. Just one of numerous examples appears on page 105 in a discussion of Van de Velde's views that the menstrual cycle caused women to suffer from emotional instability. This is followed by a comment that Naomi Wolf will later show "the price women pay for such views" in the form of thousands of breast operations and face-lifts. Such surgery was not common in Van de Velde's day. Nor was he writing about women's bodies in the section under discussion.

This example brings me to my greatest concern about the book. While I would certainly agree that marriage and sex manuals helped shape American sexual mores, it is a mistake to see only their disciplinary power. To do so denies subjectivity to those who read them. It seems likely that many women read these works, found them liberating, and knew how to select what they wanted. For example, while the majority of the advice books focused on the importance of marriage as the context for sexual relationships, increasing numbers of young women have disregarded this instruction over the course of this century. Untangling the complexities of both liberation and restriction to be found in these advice books is a more difficult task, but its attempt would have led to a richer analysis than that found here.

Queer in Russia: A Story of Self, Sex, and the Other. By Laurie Essig. Durham, N.C.: Duke University Press, 1999. Pp. xx+244. \$49.95 (cloth); \$17.95 (paper).

Annick Prieur
University of Aalborg

The cover of this book has an intriguing photo of a highly androgynous person beside the title *Queer in Russia*. I imagine this individual is "one of them." When I turn the book over, I find out the photo is of the author herself, an American sociologist. My skepticism arises: Is this a trendy and postmodern book, so inspired by the self-reflexive turn that the author fills more than the research object does? Actually, it is a trendy book, but not too trendy. The author is present in her writing but does not extend on herself. As the title indicates, the book belongs within queer

theory, but the subjects are not depicted as free to create their own lives and gender performances. The author has a deep respect for material conditions and cultural traditions and for the resulting restrictions on individual actions.

Behind the book lies comprehensive and demanding field work. With a focus on the impact for their identities, Essig writes about what the fall of the Iron Curtain has meant for subjects with a nonnormative sexual practice—subjects Westerners conceptualize as homosexual. She starts with documentation, based on interviews and written sources, of the conditions for people with a same-sex preference under the Soviet era. As historical analyses, these chapters are somewhat sketchy and uncritical, but as a background for her analysis of the present, the historical outline serves its purpose. The author gives a truly shocking description of the prison and work camp conditions a great number of men (suspected of) having had sex with men were subjected to. Not less shocking, but perhaps more surprising, is the information about the extensive use of psychiatry against women suspected of same-sex relations. To what extent this practice was interrupted by the change of regime, however, remains unclear. The most surprising information is about the rather frequent gender or sex reassignments for persons, and particularly for women, whom Westerners would have labeled gays or lesbians, but who in Russia are labeled transsexuals. For many, an official reassignment as the other gender, through a legal change of passport, has been enough to let them live in peace and even marry a person of their own sex. Quite a few have undergone sex reassignment surgery after their gender reassignment.

Unfortunately, we do not learn much about what has become of these persons. For ethical and methodological reasons, Essig has chosen to focus on public manifestations of "queerness," at the expense of private lives. I am not convinced this was a good choice, and in any case, it shows itself difficult to maintain. Indeed, she has an interesting story to tell about the difficult launch of a "gay and lesbian" movement and its short blossoming before disillusion and divisions replaced the first postperestroika optimism. Nevertheless, the stories about living subjects are the ones that I will remember the best, for instance about the couples of teenage girls who look like each other, but where one is defined as transsexual and the other as normal. Their ways of maneuvering within the strict limits of a still very repressive society, and their ways of conceiving of their selves, are indeed fascinating. The pivotal point of the book is the identity constructions, or rather, as Essig conceptualizes them in order to grasp their openness or fluidity, the *subjectivities*. (Instead of subsuming them under "postidentity politics," she could have grasped the occasion to criticize a notion that suits badly for subjects who have never known modern identity constructions. I believe they exist in many societies besides modern conceptions, remaining a lower-class phenomenon.) Essig shows that while perestroika probably has had an influence on practices and curbed the most repressive state interventions, these subjectivities do not seem to have changed so much. This might explain the

lack of a sustainable movement forming around the issue, as an activism without identity is delicate. Essig knows Russia well enough to be skeptical about Western interpretations of the Russian subjects ("closeted homosexuals") and solutions proposed to them ("coming out"), as well as to criticize the cultural imperialism of exporting identity politics. She also understands some Russian "queers'" paradoxical adherence to nationalism and fascism out of a painful opposition between Russia and the West.

It is unfair to end the review of such a rich and well-written book with criticism, but I must admit I am quite critical about the concluding chapter. Introducing self-written fiction stories about sexual encounters in Russia, the author breaks with the research genre before she indulges in a less sociological and more romanticized version of postmodernism than the one that has guided her work. Instead of illustrating the "endless possibilities of creative desires," her analysis, as I read it, in an insightful way shows the culturally specific logics behind alternative choices and combinations.

A Thousand Screenplays: The French Imagination in a Time of Crisis. By Sabine Chalvon-Demersay. Translated by Teresa Lavender Fagan. Chicago: University of Chicago Press, 1999. Pp. 200. \$30.00 (cloth); \$15.00 (paper).

S. Craig Watkins
University of Texas, Austin

Looking for fresh and entertaining material, the officials at France Television held a contest. The short-term objective, author Sabine Chalvon-Demersay tells us, "was to offer a springboard for a new wave of creators, by producing in the near future one hundred first works of young authors, 'idea people', or directors from France, Europe, or southern countries" (p. 1).

The invitation elicited a phenomenal response, as 1,120 screenplays were submitted. The screenplays came from a diverse range of writers representing various geographical regions, ages, racial and ethnic groups, and professionals and irregular workers. Chalvon-Demersay decided to subject the fictional scenarios to sociological analysis. The initial inspection of the numerous screenplays produced some intriguing patterns. She writes: "looking at all the texts, in spite of their differences, we continually found the same elements echoing back and forth: the same characters involved in the same plots, confronted with the same type of difficulties, acting according to the same principles, employing the same sort of reasoning" (p. 2). Further, she adds, "set side by side, they depicted the contours of the same universe. Clearly these authors had constructed a common world" (p. 3).

After establishing a very interesting sociological phenomenon, Chalvon-Demersay uses the remaining four chapters of the book to describe how

the screenplays construct a common world overwhelmed by crisis. First, the majority of the screenplays were set in contemporary France. As a result, many of the narratives pivot around contemporary trends, issues, and social problems. Second, many of the screenplays lament the passing of social cohesion while emphasizing feelings of powerlessness and misery. Those individuals who managed to maintain a certain degree of security and power in the modern world—developers, police, doctors—were portrayed as villains. And while beggars, immigrants, youth, and the homeless were less likely to be depicted in villainous terms, they were seldom treated as heroes either. Rather, they were often the weary victims of rapid social disorganization and dislocation.

Another finding of note was how institutions figured in most of the screenplays. Rather than depicting institutions as the locus of power, control, and order, they were constructed as powerless, ineffective, and unable to maintain meaningful experiences of social cohesion. Chalvon-Demersay notes that in many of the screenplays there is no real sense of community, common cause, or order. The screenplays point to a nation in crisis, one that shows no signs of relief or recovery. Chalvon-Demersay discusses several other parallels.

While descriptions of the similarities that connect the screenplays are necessary to substantiate her main claim, they overwhelm this small book. What is a potentially interesting phenomenon loses much of its appeal. As the reader marches through the main chapters, a number of questions come to mind. One question relates to the specter of crisis that drives many of the screenplays. The reader could be helped immensely if Chalvon-Demersay developed at least one chapter around the public discourses that seem to be gripping France at the end of the 20th century. More specifically, the reader does not get a sense of the social context from which the screenplays are created. The screenplays give more than a glimpse of the kinds of issues percolating in French society—technological change, immigration, housing, economic inequality, and the growing chasm between the suburbs and the central cities. Still, the reader never gets a sense of how these issues resonate in French society, that is, how they help define how French citizens understand and make sense of their world.

Thus *A Thousand Screenplays* does a good job of discussing the screenplays but fails to adequately consider the material world they engage. Many of the screenplays are obviously in dialogue with a larger arena of public discourse, yet they are analyzed in what amounts to a social and historical vacuum, thus diminishing the power they might otherwise achieve as important sites of discourse.

The other major question—who were the authors of the screenplays and what position in French society do they occupy?—is addressed in the appendices. In fact, the most arresting findings are reserved for the appendices that make up the back-of-the-book matter. For example, after revisiting the initial question—What makes the screenplays so revealing?—Chalvon-Demersay concludes that it is not so much the personal

histories of the authors, but their perception of the kinds of narratives, characters, and conflicts that potential audiences were most likely to derive pleasure from that connected such a varied body of French citizens. Chalvon-Demersay notes that the common world she encounters in the screenplays is all the more astounding because the authors, although scattered across various sectors of French life, believed that a certain type of fictional world would be preferred by millions of television viewers.

The implications of this finding are far more interesting and deserve more attention than the author allows. The finding suggests that how cultural producers construct the world is inextricably linked to how they think their intended or imagined audiences are most likely to experience or understand the world. Because so much of the book is devoted to analyzing the screenplays, very little space is devoted to analyzing how crisis-tinted discourses and French society are implicated in the imaginary worlds embodied in the screenplays.

If you are interested in learning about some of the bizarre and fascinating scenarios aspiring screenplay writers create to represent contemporary French society, you will probably find this book absorbing. If, however, you are interested in studying how the production of culture is linked sociologically to broader social currents, media organizations, and the creative process, you may come away disappointed.

Mission Improbable: Using Fantasy Documents to Tame Disaster. By Lee Clarke. Chicago: University of Chicago Press, 1999. Pp. xi+217.

Peter K. Manning
Michigan State University

This punnishly titled book concerns the function of fantasy documents in managing risk. A fantasy document is a tool "of persuasion designed to create the impression of expertise for certain audiences" (p. 137). The function of the fantasy document is to promote the impression that disasters, fraught with uncertainty (unknowable outcomes), produce risks (matters subject to probable outcomes). Clarke uses government documents, newspaper accounts, and published case studies of disasters to illuminate materials on the Exxon-Valdez oil spill, controversies surrounding the safety of the Shoreham Nuclear Power Plant, and civil defense preparations for nuclear war.

Planning, fantasy documents, and the ways bureaucrats and other experts have assessed and prepared for disasters occupy the first three chapters. Planning is an analogical process, according to Clarke, involving seeing some things as like others. Planning is linked in the modern state with authority, experts, and their rhetorics. In the final chapter, he explores five explanations for the production of such documents, what might be called "as ifs,"—intentional lies, psychological motives, political

palliatives, managerial imperatives, and expert knowledge. He favors the latter two and finds conflict within and between powerful agencies as the arena in which documents are created. He concludes that "our risky systems require a logic that leads to all sorts of failures. We increasingly depend on systems that have catastrophic potential" (p. 166). The bureaus and their agents "normalize danger" in large part by producing, circulating, and dramatizing misleading yet persuasive planning documents. Key audiences legitimate the agencies, which in turn create more documents.

There are a great many inferences about people, processes, and meanings here. Audiences, by and large, want to be persuaded and to legitimate experts, yet in each of these three cases, public movements, political response, federal action, and laws resulted because these documents and agencies were not believed. Civil defense no longer preoccupies the public not because the Cold War is finished, but because it is not possible, as the British Medical Association concluded some 20 years ago, to adequately respond to a nuclear attack. Shoreham is closed. The risk from the breakup of oil tankers remains. What precisely makes a document persuasive to which audiences, and when? This dramaturgical question cries out for empirical specification.

This book is in the emergent category of secondhand ethnography. These books, of which Perrow's *Normal Accidents* (Basic Books, 1984) is exemplary, use documents to mount functionalist, materialist interpretations of complex, interconnected processing of information and human deciding—deciding done over many years, in quite different social contexts, often secret, and with deep dramaturgical meaning and potential. These rather speculative works can be compared with works of Carole Heimer, Wendy Espeland, John Van Maanen, and Martha Feldman—works that depict closely decisions in organizational context.

From public reports, Clarke judges some responses and planning for disaster appropriate (chap. 3) after the fact, calling them nonfantasy documents because they contained "a higher ratio of instrumental to symbolic rationality" (p. 47). The bases for calculating this pseudopositivistic "ratio" are not divulged, and the formulation begs the question of the utility of such a binary distinction applied at a distance of many years. The line between "fantasy" and well-imagined planning documents remains blurred. Indeed, one might ask what sort of document is not intended to persuade, selectively illuminate, induce, and engage an audience?

Simmel taught us that the basis of a group is shared secrets. Clarke's research deals with secret and quasi-secret activity bearing on government security. Government secrets, even in this country, are difficult to breach, reveal, and disentangle from malice, greed, ignorance, and error. The history of government security, whether it involves energy (oil, gas, nuclear power), defense preparations, or strategic vulnerability depends in part on lies, official secrets, public deception, misinformation, and disinformation. The intent, direction, validity, and purpose of government documents in these domains is dubious and factious. Government re-

quires systematic lying in the interests of its own perpetuation, and it makes secret that which threatens its interests.

Clarke dismisses "talk about talk" as a basis for theory; seeing it as essentially a fallacious idealist conceit, and follows his conceit assiduously. Unlike Diane Vaughan's imaginative reconstruction of deciding based on archival records and participant interviews (*The Challenger Launch Decision* [University of Chicago Press, 1996]), this book is arm-chair sociology.

Risk and fantasy are not always tied together so neatly. It is notable to consider the current public fascination of the media and intellectuals with risk, increasingly simulated, commodified, and distant from actual face-to-face threats. The progress of civility comes not through rationalization, but through changing configurations, shared collective actions—from brutal face-to-face contests, organized games, and competitions, to spectacles and enchantments (theme parks, historical simulations, televised wars). These created forms are accompanied by the occasional explosive burst of excess—hurricanes, floods, tornadoes—worldwide and reported almost instantly across the world by CNN. People in the Anglo-American world live in the most benign of times, with low risk of death, long life expectancy, and a fascination with exotic risk, some of it vicarious, like watching missiles destined for Baghdad, entirely without personal danger. We also endure a family of texts that exploit and amplify other people's risks—gossip, the Starr Report, books on Monica Lewinsky, analysis of the O.J. trial, and the several presently popular high-risk narratives on death via storms, mountain climbing, and yacht racing. These texts, like natural events and disasters, provide counterpoint to the rationalizing world around us. Furthermore, they suggest that control, "taming," or "managing" risk is in some dialectic with seeking it—in the Himalayas, at sea, on motocross bikes, skateboards, and hanging from sheer rock walls. Is it necessary to "persuade" people that life is both risky and uncertain? How are these public documents related to the everyday world of deciding?

Problems of Form. Edited by Dirk Baecker. Stanford, Calif.: Stanford University Press, 1999. Pp. xii+248. \$55.00 (cloth); \$22.95 (paper).

Stephan Fuchs
University of Virginia

Problems of Form collects contributions from students, associates, and fellow travelers of the late Niklas Luhmann. Its focus is on what the calculus of the British mathematician Spencer-Brown can do for sociology, specifically the theory of social systems, which largely remains Luhmann's brainchild. He has remodeled that theory several times; the latest transformation before his death in 1998 was to practice sociology as second-order observing of observers. Society is composed of observers

who communicate; sociology observes how social observers, including itself, observe; what they can and cannot see, and how the semantics of observation change over time.

Spencer-Brown's "calculus of indication" is the mathematical foundation of systems theory. The main rule of the calculus is to draw a distinction. Nothing can be observed without distinctions. The "form" of the distinction is everything it does. A distinction marks something in the world—*this*, not that. The "that" is the not-*this*, that is, the state of the world unmarked by the distinction. The world now has a marked and an unmarked state, but it also gains the distinction itself, including the observer who draws it in *this* way rather than another.

Distinctions are contingent. They can be drawn or not and they can be drawn in different ways. Observers differ according to how they draw distinctions. Since observers are part of the world, and since observing is an operation that either happens or does not, observers and observations can be observed themselves. This is the level of sociology, or second-order observing.

Societies differ in how many observers they accommodate. Highly stratified societies restrict legitimate or official observing to the top: a palace, church, or center. These are distinctions that appear not contingent but necessary, as in tune with the essence of the world. Authoritative distinctions hide contingency in taboo, secrecy, or sacred texts. Modern society has very many competitive observers who observe each other also, an act that generates ideological politics. The suspicion grows that observers have hidden agendas, and the critical task of sociological observing is to unmask ideology. This is the intellectual thrust of much Marxism.

Systems theory is radically constructivist—including about itself—as one observer among others. As society loses its center or apex, the possibilities for "privileged" representations—observations that are more than observations—decline. What is left is a pluralism of many observers, conflicts over who is and is not an observer, and struggles over what an observer can and cannot observe—with the distinction or distinctions that those observers are observed to employ. "Truth" becomes elusive, since it no longer has any metaphysical guarantees and safeguards. Instead, "truth" becomes an internal accomplishment of recursively coupled observations. We might call such networks of observation "cultures."

What can be done with this unusual and fresh idea? I will pick three examples. Japp analyzes the form of the new social protest movements. They protest against society within society. To obscure this paradox, the form of the distinction is moralized so that "protest" is good and "society" bad. Protest movements tend to coalesce around low probability/high impact disasters, since those cannot be handled by politics as usual. Since such disasters remain always possible, the movement has enough worries to keep stimulating itself into protest activities. It is not that movements are generated by an issue; rather, the movement emerges first and then seizes the issues that make its own continuance more likely. Internally, the movement creates solidarity and righteousness as well as an expres-

sive state of "being concerned" (*Betroffenheit*). The movement members are concerned—not about anything in particular, since such problems might be solved and go away, but about society or culture "in general." Hence, movements generalize with abandon—the enemy is men, the ruling class, multis, or instrumental reason.

Baecker uses Spencer-Brown's "reentry" to analyze play as socialness. When distinctions, such as play/nonplay, reenter the domain of what they distinguish, play observes itself and is being observed as play, not as not-play. Reentry generates contingency on a second-level order of observation. In this way, operations turn into distinctions, that is, contingent incisions into the world, performed by one among many observers. These observers can be observed by different observers. An observer who plays (or not?) with the distinction play/nonplay celebrates sheer contingency.

In the most somber chapter, Corsi brings out the dark side of careers by distinguishing them from aging. Careers are what persons without true destinies have instead. One must make them and must have the cooperation of others to do this. Careers can be chosen; destinies are realized and fulfilled or not. With age, however, come certain restrictions on the moves one can still anticipate to make in the future. This is when the contingency of career encounters the necessity of fate and destiny. Not much time is left; the options narrow down, and the future possibilities become limited by death. The self-made man is now at a loss. Readers interested in Luhmann's work should consult this book, but are also encouraged to go back to his own writings.

AJS

American Journal of Sociology

Volume 105 Bimonthly July 1999–May 2000

ROGER V. GOULD, Editor

EMILY BARMAN, JIM DAVIS, ANDREAS GLAESER, TONY PAIK, JENNIFER PASHUP, JESPER SØRENSEN, TOBY STUART, and GENEVIÈVE ZUBRZYCKI, Associate Editors

LESLIE SALZINGER, Associate Editor, Book Review

JENNIFER TELLO BUNTIN, KELLY DALEY, DAVID GRAZIAN, JULIA GWYNNE, SUKARI IVESTER, HARRIS KIM, CHI-YOUNG KOH, KWAI HANG NG, and STEPHEN ROSENBERG, Book Review Editorial Board

SUSAN ALLAN, Managing Editor

CONNIE FRITSCHÉ and GRETCHEN KALWINSKI, Editorial Assistants

Consulting Editors VICTORIA D. ALEXANDER • CHRISTOPHER ANSELL • NICOLA BEISEL • YANJIE BIAN • WILLIAM BRUSTEIN • STEVEN CLAYMAN • SAMUEL COHN • FRANK DOBBIN • SCOTT ELIASON • MUSTAFA EMIRBAYER • WENDY NELSON ESPELAND • MYRA MARX FERREE • JACK GOLDSTONE • PHILIP GORSKI • GUANG GUO • LINGXIN HAO • KAREN HEGTVEDT • BARBARA KILBOURNE • RUUD KOOPMANS • JOHN LAUB • EDWARD J. LAWLER • MICHAEL LOVAGLIA • SAMUEL R. LUCAS • MICHAEL MACY • ROSS MATSUEDA • DEBRA C. MINKOFF • CAROL MUELLER • JOANE P. NAGEL • HAYAGREEVA RAO • WILLIAM G. ROY • DIETRICH RUESCHEMEYER • JANET CHAFETZ SALTZMAN • PAMELA SMOCK • SCOTT SOUTH • MARK STEINBERG • TONY TAM • BRIAN UZZI • BRUCE WESTERN • MICHAEL WHITE • RAYMOND WONG

Editorial Board ANDREW D. ABBOTT • GARY S. BECKER • CHARLES E. BIDWELL • DONALD J. BOGUE • ANTHONY BRYK • RONALD BURT • TERRY N. CLARK • JOHN L. COMAROFF • LARRY HEDGES • PATRICK HEUVELINE • EVELYN KITAGAWA • EDWARD O. LAUMANN • DONALD N. LEVINE • WILLIAM L. PARISH • MARTIN RIESEBRODT • ROBERT J. SAMPSON • SASKIA SASSEN • ROSS M. STOLZENBERG • FRED L. STRODTBECK • GERALD D. SUTTLES • RICHARD TAUB • LINDA J. WAITE • KIM WEEDEN • KAZUO YAMAGUCHI • DINGXIN ZHAO

THE UNIVERSITY OF CHICAGO PRESS
CHICAGO, ILLINOIS

© 1999, 2000 by The University of Chicago
All rights reserved
Published July, September, November 1999
January, March, May 2000
CHICAGO, ILLINOIS, USA
PUBLISHED BY THE UNIVERSITY OF CHICAGO PRESS

ACKNOWLEDGMENTS TO REFEREES

The editors are indebted to their many colleagues who assist them in the assessment of manuscripts. The work of these referees is a continuous service to the profession. Following is a list of people who read at least one manuscript in 1999. We take this opportunity at the end of volume 105 to thank them.

PETER ABELL	KENNETH BOLLEN	ELISABETH CLEMENS
MITCHEL Y. ABOLAFIA	PHILLIP BONACICH	RAM CNAAN
JULIA P. ADAMS	TERRY BOSWELL	IRA J. COHEN
PATRICIA A. ADLER	ANNE BOWLER	LAWRENCE COHEN
PETER ADLER	MARY ELLEN BOYLE	PATRICIA CLINE COHEN
ARI ADUT	JOHN BRAITHWAITE	H. M. COLLINS
OMAR ALTALIB	NEIL BRENNER	RANDALL COLLINS
ROBERT ALTHAUSER	DANIEL BRESLAU	PAUL COLOMY
EDWIN AMENTA	KARIN L. BREWSTER	ROBERT WILLIAM
CAROL ANESHENSEL	WILLIAM P. BRIDGES	CONNELL
ROBERT J. ANTONIO	JOHN BRIGHAM	STEVE COOK
MASAHIKO AOKI	JULIE BRINES	MARK COONEY
BRANDON APPELEGATE	MARY C. BRINTON	AMY CORNING
WILLIAM G. AXINN	CLIFFORD L. BROMAN	SHELLEY J. CORRELL
SARAH L. BABB	CHARLES BROUGHTON	JEANETTE COVINGTON
KIM BABON	CHRISTOPHER R.	DIANA CRANE
BRIAN ALAN BAKER	BROWNING	ROBERT CROSNOE
DAVID BAKER	ROGERS BRUBAKER	KYLE D. CROWDER
WAYNE E. BAKER	STEVE BRUCE	CATHERINE RIEGLE-
KAREN BARKEY	VINCENT BUSKENS	CRUMB
STEPHEN R. BARLEY	JON BUTLER	JAMES CUMMINS
EMILY BARMAN	ALBERTO CAMBROSIO	TIMOTHY CURRY
WILLIAM P. BARNETT	CHARLES CAMIC	JAMES W. AINSWORTH-
DAVID BARRON	CAMERON CAMPBELL	DARNELL
JOHN P. BARTKOWSKI	KAREN E. CAMPBELL	DIANE DAVIS
ANNE BARTLETT	MARIA CANCIAN	GERALD F. DAVIS
PETER S. BEARMAN	PATRICK J. CARR	MURRAY S. DAVIS
E. M. BECK	GLENN R. CARROLL	SETH DEAVILA
PENNY EDGELL BECKER	BRUCE G. CARRUTHERS	MATHIEU DEFLEM
MARK BEISSINGER	LYNNE M. CASPER	N. J. DEMERATH
MICHAEL M. BELL	KARIN D. KNORR-	FANG DENG
MABEL BEREZIN	CETINA	MARIO DIANI
ANNETTE BERNHARDT	DANIEL F. CHAMBLISS	THOMAS DIETZ
MARY BERNSTEIN	WILLIAM J. CHAMBLISS	PAUL DiMAGGIO
DIANE BESSEL	KUANG-CHI CHANG	THOMAS A. DiPRETE
KRAIG BEYERLEIN	CAMILLE ZUBRINSKY	JO DIXON
SUZANNE M. BIANCHI	CHARLES	TIMOTHY J. DOWD
DENISE BIELBY	MARIA CHARLES	DANIEL DREZNER
ELISA J. BIENENSTOCK	MARK CHAVES	JULES DUCHASTEL
DWIGHT B. BILLINGS	KEVIN CHEN	GREG J. DUNCAN
FRED L. BLOCK	THEODORE CHIRICOS	RACHEL E. DWYER
HANS-PETER BLOSS-	SAMUEL CLARK	STEPHEN J. ELLINGSON
FELD	TERRY CLARK	CHRISTOPHER G.
JACQUELINE BOLES	DAN CLAWSON	ELLISON

DON EMMERSON
 PAULA ENGLAND
 CHARLES R. EPP
 STEVEN EPSTEIN
 JOHN H. EVANS
 PETER B. EVANS
 GEORGE FARKAS
 REYNOLDS FARLEY
 ROBERT R. FAULKNER
 KATHERINE FAUST
 ADRIAN FAVELL
 SCOTT FELD
 DIANE FELMLEE
 MARY FENNELL
 PRISCILLA P.
 FERGUSON
 ROBERTO M.
 FERNANDEZ
 KENNETH F. FERRARO
 MYRA MARX FERREE
 GARY ALAN FINE
 GLENN FIREBAUGH
 NEIL FLIGSTEIN
 ERIC FONG
 SIGNITHIA FORDHAM
 DAVID JOHN FRANK
 KENNETH FRANK
 ROBERT F. FREELAND
 LINTON C. FREEMAN
 R. SCOTT FREY
 NOAH E. FRIEDKIN
 HARRIET FRIEDMANN
 CAROLYN FUQUA
 M. ELIZABETH FUSSELL
 JOSEPH GALASKIEWICZ
 SALLY K. GALLAGHER
 DIEGO GAMBETTA
 WILLIAM A. GAMSON
 VENELIN I. GANEV
 ROSEMARY GARTNER
 VIKTOR GECAS
 CHARLES GEISLER
 THEODORE P. GERBER
 GARY GEREFFI
 KATHLEEN GERSON
 DAVID GIBSON
 BERNHARD GIESEN
 A. R. GILLIS
 MARY GILLMORE
 MARCO G. GIUGNI
 ANDREAS GLAESER

JENNIFER GLASS
 NORVAL D. GLENN
 WALTER GOLDFRANK
 FRANCES
 GOLDSCHIEDER
 ERICH GOODE
 JEFF GOODWIN
 LINDA GORDON
 PAUL DE GRAAF
 EDWARD GRABB
 DONALD GRANBERG
 RICK GRANNIS
 MARK GRANOVETTER
 HAROLD G. GRASMICK
 DAVID GRAZIAN
 DAVID GREATBATCH
 DONALD GREEN
 DAVID F. GREENBERG
 HENRICH R. GREVE
 ERIC GRODSKY
 MAURO F. GUILLEN
 DOUGLAS J. GUTHRIE
 JULIA GWYNNE
 DAVID S. HACHEN
 JOHN HAGAN
 JOHN HAGEDORN
 WARREN HAGSTROM
 JOHN R. HALL
 MAUREEN T.
 HALLINAN
 GARY HAMILTON
 SHIN-KAP HAN
 BLACK HAWK HANCOCK
 MICHAEL T. HANNAN
 HEATHER HAVEMAN
 JEFFREY HAYDU
 JEFFREY HAYES
 MOSHE HAZANI
 DOUGLAS D. HECKA-
 THORN
 DAVID M. HEER
 KAREN HEIMER
 DAVID R. HEISE
 RONALD HELMS
 JOSEPH C. HERMANO-
 WICZ
 GARY HERRIGEL
 ALEXANDER HICKS
 PAUL M. HIRSCH
 CHARLES HIRSCHMAN
 JENNIFER HOCHSCHILD

JOHN P. HOFFMAN
 RUTH HOROWITZ
 MICHAEL HOUT
 MIN-HSIUNG HUANG
 ROBERT HUMMER
 SCOTT A. HUNT
 PAUL INGRAM
 JOSE ITZILGSOHN
 DAVID JACOBS
 RONALD N. JACOBS
 DAVID R. JAMES
 THOMAS JANOSKI
 PAUL A. JARGOWSKY
 GUILLERMINA JASSO
 J. CRAIG JENKINS
 KENT JENNINGS
 DAPHNE JOHN
 JENNIFER JOHNSON
 ROBERT EMMET JONES
 LUTZ KAELEBER
 ARNE L. KALLEBERG
 JAAP KAMPS
 SATOSHI KANAZAWA
 MICHAEL A. KARESH
 ELIHU KATZ
 JACK KATZ
 ROBERT L. KAUFMAN
 MARIA J. KEFALAS
 LISA A. KEISTER
 M. VERA KEITH
 JOHN KENNY
 HOWARD KIMELDORF
 ANTHONY KING
 BRETT KLOPP
 STAN KNAPP
 MARY ELLEN
 KONIECZNY
 ROBERTO PATRICIO
 KORZENIEWICZ
 ERIC C. KOSTELLO
 JOHN HENRY KRAMER
 GARY A. KREPS
 GRETA R. KRIPPNER
 CANDACE KRUTT-
 SCHNITT
 RICHARD W. LACH-
 MANN
 KENNETH C. LAND
 NANCY LANDALE
 RICK LARRICK
 JANET LAURITSEN

Acknowledgments

ROBIN LEIDNER
ERIC M. LEIFER
CHARLES LEMERT
ROBERT I. LERMAN
ANDREA LEVERENTZ
DANIEL T. LICHTER
IVAN LIGHT
NAN LIN
BRUCE G. LINK
MICHAEL LIPSKY
LYN H. LOFLAND
JOHN ALLEN LOGAN
JOHN R. LOGAN
MARY BLAIR-LOY
STANFORD M. LYMAN
THOMAS A. LYSON
JOHN VAN MAANEN
HEATHER MACINDOE
GERALD MACKIE
DAG MACLEOD
MICHAEL W. MACY
RICHARD P. MADSEN
JENNA W. MAHAY
PETER K. MANNING
WENDY D. MANNING
ELAINE MARCHENA
NOAH MARK
HELEN M. MARKS
GERALD MARWELL
DOUGLAS S. MASSEY
ROSS L. MATSUEDA
REUBEN A. B. MAY
DOUGLAS W. MAYNARD
LESLIE MCCALL
JOHN D. MCCARTHY
JAMES M. MCCORMICK
ALLAN MCCUTCHEON
DANIEL MCFARLAND
PATRICK MCGUIRE
PAUL MCLEAN
J. MILLER MCPHERSON
ELIZABETH
MCSWEENEY
DAVID MECHANIC
ELIZABETH MENAGHAN
STEVEN F. MESSNER
STJEPAN G. MESTROVIC
DAVID S. MEYER
STEPHEN J. MEZIAS
NEAL MILNER
PYONG GAP MIN

BETH A. MINTZ
RICHARD MITCHELL
MARK S. MIZRUCHI
SUZANNE MODEL
PHYLLIS MOEN
JOHN MOHR
LINDA D. MOLM
HARVEY MOLOTCH
JAMES MONTGOMERY
DAWNE MOON
GWENDOLYN MOORE
KELLY MOORE
HARRIET P. MORGAN
LAURIE MORGAN
LESLIE MORGAN
ZIAD MUNSON
VICTOR G. NEE
KWAI HANG NG
FRANÇOIS NIELSEN
LUIS CARLOS ROBERTO
NOGALES
GUY OAKES
DENIS O'HEARN
CHRISTINE OLIVER
PAMELA OLIVER
MICHAEL OLNECK
SUSAN OLZAK
KARL-DIETER OPP
MARYJANE OSA
SUSAN OSTRANDER
ELINOR OSTROM
ROBIN WAGNER-
PACIFICI
RON PAGNUCCO
FRED C. PAMPSEL
TOBY PARCELL
SILVANA PATRIARCA
MARY E. PATTILLO-
MCCOY
PAMELA PAXTON
NICHOLAS PEDRIANA
BERNICE A. PESCO-
SOLIDO
TROND PETERSEN
STEPHEN PETTERSON
STEVEN PFAFF
MAX J. PFEFFER
DAMON JEREMY PHIL-
LIPS
ANDREW PICKERING
JANE ALLYN PILLAVIN

TREVOR PINCH
FRANCES FOX PIVEN
STEPHEN B. PLANK
JOEL M. PODOLNY
DAVID POPENOE
PAMELA A. POPIELARZ
DONATELLA DELLA
PORTA
DANIEL POWERS
MONICA PRASAD
LINDSAY PRIOR
ZHENCHAO QIAN
JILL QUADAGNO
LINCOLN QUILLIAN
PAMELA QUIROZ
CHARLES RAGIN
ERIC RAMBO
BRUCE H. RANKIN
KEVIN RATHUNDE
WERNER RAUB
RAY E. REAGANS
MICHAEL J. REAY
ANISA C. RHEA
CECILIA RIDGEWAY
RACHEL RINALDO
JAN RIVKIN
KEITH ROBERTS
ROLAND ROBERTSON
JOHN ROBINSON
JOE RODGERS
FABIO ROJAS
ELAINE ROMANELLI
EUGENE ROSA
SONYA O. ROSE
STEPHEN D. ROSEN-
BERG
MICHAEL J. ROSEN-
FELD
RACHEL A. ROSEN-
FELD
SARAH ROSENFIELD
NAOMI ROSENTHAL
WILLIAM G. ROY
DIETER RUCHT
MARTIN RUEF
STEVE L. RYTINA
SCOTT D. SAGAN
ARTHUR SAKAMOTO
ROBERT J. SAMPSON
GARY D. SANDEFUR
REBECCA L. SANDEFUR

- | | | |
|-----------------------|---------------------|---------------------|
| MARK MAURICE | GLENN A. SPITZE | SUDHIR A. VENKATESH |
| SANFORD | DAVID STARK | SIDNEY VERBA |
| JOACHIM J. SAVELSBERG | TRACY STEELE | WAYNE VILLEMEZ |
| THOMAS J. SCHEFF | CLAIRE STERK | KIM VOSS |
| KATHRYN S. SCHILLER | JAN E. STETS | ROBERT L. WAGMILLER |
| DEBRA SCHLEEF | MITCHELL STEVENS | DAVID G. WAGNER |
| WILLIAM SCHMIDT | ARTHUR STINCHCOMBE | ANDREW G. WALDER |
| MARC SCHNEIBERG | YANICK ST. JEAN | HORNG-LUEN WANG |
| BARBARA SCHNEIDER | CRAIG A. ST. JOHN | R. STEPHEN WARNER |
| ROBERT SCHOEN | JEAN STOCKARD | JOHN ROBERT WARREN |
| ANDREW SCHRANK | WOLFGANG STREECK | DUNCAN WATTS |
| MICHAEL SCHUDSON | ROBIN STRYKER | DAVID L. WEAKLIEM |
| HOWARD SCHUMAN | MARK C. SUCHMAN | KIM WEEDEN |
| BARRY SCHWARTZ | CLEA SUCOFF | SIMON E. WEFER |
| MICHAEL SCHWARTZ | JOHN R. SUTTON | BARRY WELLMAN |
| GERALD W. SCULLY | DUANE SWANK | DAVID WELLMAN |
| KAREN SECCOMBE | CHRISTOPHER | MING WEN |
| GAY SEIDMAN | SWANSON | ELIZABETH WEST |
| STEVEN SEIDMAN | RICHARD SWEDBERG | BRUCE WESTERN |
| PHILIP SELZNICK | KARL E. TAEUBER | AMY S. WHARTON |
| GERSHON SHAFIR | ILAN TALMUD | DOUGLAS R. WHOLEY |
| MICHAEL J. SHANAHAN | JOSEPH B. TAMNEY | MARTIN K. WHYTE |
| SUZANNE SHANAHAN | MARYLEE C. TAYLOR | WILLIAM B. WILCOX |
| KIMBERLEE A. | VERTA TAYLOR | MELISSA WILDE |
| SHAUMAN | JAY D. TEACHMAN | CHRISTINE WILLIAMS |
| BETH ANNE SHELTON | GEORGE THOMAS | DAVID R. WILLIAMS |
| VLADIMIR SHLAPEN- | MICHAEL C. | RHYS H. WILLIAMS |
| TOKH | THORNTON | JOHN WILLIAMSON |
| EDUARDO BONILLA- | MARTA TIENDA | KENNETH L. WILSON |
| SILVA | CHARLES TILLY | CHRISTOPHER WINSHIP |
| RITA J. SIMON | JEFFREY M. | ROGER WOJTKIEWICZ |
| IDA HARPER SIMPSON | TIMBERLAKE | RACHAEL ANNE |
| MILES SIMPSON | EDWARD A. TIRYAKIAN | WOLDOFF |
| JOHN SKVORETZ | MIKK TITMA | ERIK OLIN WRIGHT |
| JOHN SMAIL | CHARLES R. TITTLE | JOLYON WURR |
| NEIL SMELSER | PAMELA S. TOLBERT | YU XIE |
| CHRISTIAN SMITH | STEWART E. TOLNAY | VALERY YAKUBOVICH |
| DAVID HORTON SMITH | DOLORES TREVIZO | TOSHIO YAMAGISHI |
| JACKIE SMITH | LISA TROYER | WILLIAM YANCEY |
| SANDRA S. SMITH | SHERRY R. TURKLE | JOHN MILTON YINGER |
| TOM SMITH | JONATHAN H. TURNER | YUVAL YONAY |
| MARGARET R. SOMERS | PETER UHLENBERG | YOOSIK YOUM |
| AAGE SØRENSEN | JEFFERY T. ULMER | C. WESLEY YOUNTS |
| ANNEMETTE SØRENSEN | WOUT ULTEE | MAYER N. ZALD |
| OLAV SØRENSEN | DAWN UPCHURCH | DINGXIN ZHAO |
| CALEB SOUTHWORTH | BERT USEEM | MIN ZHOU |
| DAPHNE SPAIN | THOMAS W. VALENTE | XUEGUANG ZHOU |
| KENNETH SPENNER | LINDA R. VALLI | GENEVÈVE ZUBRZYCKI |
| LYN SPILLMAN | WILLIAM VELEZ | EZRA ZUCKERMAN |

CONTENTS OF VOLUME 105

Articles

- 1249 BAUMANN, SHYON. See LIEBERSON
- 321 BIBLARZ, TIMOTHY J., and ADRIAN E. RAFTERY. Family Structure, Educational Attainment, and Socioeconomic Success: Rethinking the "Pathology of Matriarchy"
- 1175 CAO, YANG, and VICTOR G. NEE. Comment: Controversies and Evidence in the Market Transition Debate
- 1288 CASTILLA, EMILIO J. See FERNANDEZ
- 1658 CHANG, MARIKO LIN. The Evolution of Sex Segregation Regimes
- 128 CHWE, MICHAEL SUK-YOUNG. Structure and Strategy in Collective Action
- 913 CORNING, AMY. See SCHUMAN
- 1063 CRESS, DANIEL M., and DAVID A. SNOW. The Outcomes of Homeless Mobilization: The Influence of Organization, Disruption, Political Mediation, and Framing
- 455 DOBBIN, FRANK. See KELLY
- 1249 DUMAIS, SUSAN. See LIEBERSON
- 406 EDELMAN, LAUREN B., CHRISTOPHER UGGEN, and HOWARD S. ERLANGER. The Endogeneity of Legal Regulation: Grievance Procedures as Rational Myth
- 1357 ENGEN, RODNEY L., and SARA STEEN. The Power to Punish: Discretion and Sentencing Reform in the War on Drugs
- 406 ERLANGER, HOWARD S. See EDELMAN
- 1288 FERNANDEZ, ROBERTO M., EMILIO J. CASTILLA, and PAUL MOORE. Social Capital at Work: Networks and Employment at a Phone Center
- 366 FOSTER, JOHN BELLAMY. Marx's Theory of Metabolic Rift: Classical Foundations for Environmental Sociology
- 1003 GANZ, MARSHALL. Resources and Resourcefulness: Strategic Capacity in the Unionization of California Agriculture, 1959-1966
- 1572 GOLDTHORPE, JOHN H. Rent, Class Conflict, and Class Structure: A Commentary on Sørensen
- 1428 GORSKI, PHILIP S. The Mosaic Moment: An Early Modernist Critique of Modernist Theories of Nationalism

- 191 HAN, SHIN-KAP, and PHYLLIS MOEN. Clocking Out: Temporal Patterning of Retirement
- 1633 HEALY, KIERAN. Embedded Altruism: Blood Collection Regimes and the European Union's Donor Population
- 157 JACOBS, DAVID, and KATHERINE WOOD. Interracial Conflict and Interracial Homicide: Do Political and Economic Rivalries Explain White Killings of Blacks or Black Killings of Whites?
- 455 KELLY, ERIN, and FRANK DOBBIN. Civil Rights Law at Work: Sex Discrimination and the Rise of Maternity Leave Policies
- 652 KOOPMANS, RUUD, and PAUL STATHAM. Challenging the Liberal Nation-State? Postnationalism, Multiculturalism, and the Collective Claims Making of Migrants and Ethnic Minorities in Britain and Germany
- 1249 LIEBERSON, STANLEY, SUSAN DUMAIS, and SHYON BAUMANN. The Instability of Androgynous Names: The Symbolic Maintenance of Gender Boundaries
- 1583 MAHONEY, JAMES. See RUESCHEMEYER
- 191 MOEN, PHYLLIS. See HAN
- 1396 MOLM, LINDA D., NOBUYUKI TAKAHASHI, and GRETCHEN PETERSON. Risk and Trust in Social Exchange: An Experimental Test of a Classical Proposition
- 1288 MOORE, PAUL. See FERNANDEZ
- 697 MUELLER, CAROL. Escape from the GDR, 1961–1989: Hybrid Exit Repertoires in a Disintegrating Leninist Regime
- 38 MYERS, DANIEL J. See OLIVER
- 1175 NEE, VICTOR G. See CAO
- 801 OCASIO, WILLIAM. See THORNTON
- 38 OLIVER, PAMELA E., and DANIEL J. MYERS. How Events Enter the Public Sphere: Conflict, Location, and Sponsorship in Local Newspaper Coverage of Public Events
- 781 PAIGE, JEFFREY M. Conjuncture, Comparison, and Conditional Theory in Macrosocial Inquiry
- 88 PAXTON, PAMELA. Is Social Capital Declining in the United States? A Multiple Indicator Assessment
- 1396 PETERSON, GRETCHEN. See MOLM
- 1 QUILLIAN, LINCOLN. Migration Patterns and the Growth of High-Poverty Neighborhoods, 1970–1990
- 1824

- 321 RAFTERY, ADRIAN E. See BIBLARZ
- 603 RAUDENBUSH, STEPHEN W. See SAMPSON
- 1583 RUESCHEMEYER, DIETRICH, and JAMES MAHONEY. A Neo-Utilitarian Theory of Class?
- 603 SAMPSON, ROBERT J., and STEPHEN W. RAUDENBUSH. Systematic Social Observation of Public Spaces: A New Look at Disorder in Urban Neighborhoods
- 913 SCHUMAN, HOWARD, and AMY D. CORNING. Collective Knowledge of Public Events: The Soviet Era from the Great Purge to Glasnost
- 1063 SNOW, DAVID A. See CRESS
- 1523 SØRENSEN, AAGE B. Toward a Sounder Basis for Class Analysis
- 652 STATHAM, PAUL. See KOOPMANS
- 1357 STEEN, SARA. See ENGEN
- 736 STEINBERG, MARC W. The Talk and Back Talk of Collective Action: A Dialogic Analysis of Repertoires of Discourse among Nineteenth-Century English Cotton Spinners
- 1105 TAKAHASHI, NOBUYUKI. The Emergence of Generalized Exchange
- 1396 TAKAHASHI, NOBUYUKI. See MOLM
- 801 THORNTON, PATRICIA H., and WILLIAM OCASIO. Institutional Logics and the Historical Contingency of Power in Organizations: Executive Succession in the Higher Education Publishing Industry, 1958–1990
- 406 UGGEN, CHRISTOPHER. See EDELMAN
- 493 WATTS, DUNCAN J. Networks, Dynamics, and the Small-World Phenomenon
- 157 WOOD, KATHERINE. See JACOBS
- 1559 WRIGHT, ERIK OLIN. Class, Exploitation, and Economic Rents: Reflections on Sørensen's "Sounder Basis"
- 957 WRIGHT, ERIK OLIN. Working-Class Power, Capitalist-Class Interests, and Class Compromise
- 1702 YAMAGUCHI, KAZUO. Multinomial Logit Latent-Class Regression Models: An Analysis of the Predictors of Gender-Role Attitudes among Japanese Women
- 1592 ZHAO, DINGXIN. State-Society Relations and the Discourses and Activities of the 1989 Beijing Student Movement

- 1135 ZHOU, XUEGUANG. Economic Transformation and Income Inequality in Urban China: Evidence from Panel Data
- 1189 ZHOU, XUEGUANG. Reply: Beyond the Debate and toward Substantive Institutional Analysis

Commentary and Debate

- 1741 COTTER, DAVID A. See ENGLAND
- 1741 ENGLAND, PAULA, JOAN M. HERMSEN, and DAVID A. COTTER. The Devaluation of Women's Work: A Comment on Tam
- 1741 HERMSEN, JOAN M. See ENGLAND
- 1752 TAM, TONY. Occupational Wage Inequality and Devaluation: A Cautionary Tale of Measurement Error

Book Reviews

- 1761 ABBOTT, ANDREW. *Department and Discipline: Chicago Sociology at One Hundred*. Arthur L. Stinchcombe
- 583 ADAMS, REBECCA G., and GRAHAM ALLAN. *Placing Friendship in Context*. Maureen T. Hallinan
- 598 AINSWORTH-VAUGHN, NANCY. *Claiming Power in Doctor-Patient Talk*. Douglas W. Maynard
- 283 AKERS, RONALD L. *Social Learning and Social Structure: A General Theory of Crime and Deviance*. Michael R. Gottfredson
- 258 ALLEN, N. J., W. S. F. PICKERING, and W. WATTS MILLER. *On Durkheim's Elementary Forms of Religious Life*. Edward A. Tiryakian
- 277 ANYON, JEAN. *Ghetto Schooling: A Political Economy of Urban Educational Reform*. John A. Kovach
- 590 APOLITO, PAOLO, ed., WILLIAM A. CHRISTIAN, JR., trans. *Apparitions of the Madonna at Oliveto Citra: Local Visions and Cosmic Drama*. Sandra L. Zimdars-Swartz
- 883 ARDITI, JORGE. *A Genealogy of Manners: Transformations of Social Relations in France and England from the Fourteenth to the Eighteenth Century*. Peter Burke
- 1814 BAECKER, DIRK, ed. *Problems of Form*. Stephan Fuchs
- 1239 BAILEY, ROBERT W. *Gay Politics, Urban Politics: Identity and Economics in the Urban Setting*. Rhys H. Williams
- 1826

- 243 BAILEY, VICTOR. *"This Rash Act": Suicide across the Life Cycle in the Victorian City*. Kevin Breault
- 252 BALKIN, J. M. *Cultural Software: A Theory of Ideology*. Susan Silbey
- 1805 BANET-WEISER, SARA. *The Most Beautiful Girl in the World: Beauty Pageants and National Identity*. Maxine Craig
- 1770 BARKER, JAMES R. *The Discipline of Teamwork: Participation and Concertive Control*. Randy Hodson
- 581 BARNES, GILL GORELL, PAUL THOMPSON, G'WYN DANIEL, and NATASHA BURCHARDT. *Growing Up in Stepfamilies*. Linda J. Waite
- 1246 BASZANGER, ISABELLE. *Inventing Pain Medicine: From the Laboratory to the Clinic*. Nicholas A. Christakis
- 531 BATES, ROBERT H., AVNER GREIF, MARGARET LEVI, JEAN-LAURENT ROSENTHAL, and BARRY R. WEINGAST. *Analytic Narratives*. Jack A. Goldstone
- 1799 BECKER, PENNY EDGELL. *Congregations in Conflict: Cultural Models of Local Religious Life*. Mitchell L. Stevens
- 1488 BERDAHL, DAPHNE. *Where the World Ended: Re-unification and Identity in the German Borderland*. Andreas Glaeser
- 880 BJORKLUND, DIANE. *Interpreting the Self: Two Hundred Years of American Autobiography*. David R. Maines
- 870 BLAND, LUCY, and LAURA DOAN, eds. *Sexology in Culture: Labelling Bodies and Desires and Sexology Uncensored: The Documents of Sexual Science*. John H. Gagnon
- 1234 BLUM, LINDA M. *At the Breast: Ideologies of Breastfeeding and Motherhood in the Contemporary United States*. Barbara Katz Rothman
- 909 BOIX, CHARLES. *Political Parties, Growth and Equality: Conservative and Social Democratic Economic Strategies in the World Economy*. Thomas Janoski
- 859 BRADLEY, HARRIET. *Gender and Power in the Workplace: Analysing the Impact of Economic Change*. Sylvia Walby
- 1780 BRAH, AVTAR, MARY J. HICKMAN, and MÁIRTÍN MAC AN GHAIL, eds. *Global Futures: Migration, Environment and Globalization*. Peter J. Taylor
- 562 BRODKIN, KAREN. *How Jews Became White Folks and What That Says about Race in America*. Ruth Frankenberg

- 261 BROMLEY, DAVID G., ed. *The Politics of Religious Apostasy: The Role of Apostates in the Transformation of Religious Movements*. William Sims Bainbridge
- 1790 BROWN, MICHAEL K. *Race, Money, and the American Welfare State*. Dalton Conley
- 844 BRUBAKER, ROGERS. *Nationalism Reframed: Nationhood and the National Question in the New Europe*. E. J. Hobsbawm
- 1510 BRUNO, ROBERT. *Steelworker Alley: How Class Works in Youngstown*. Charles S. Varano
- 275 BRYK, ANTHONY S., PENNY BENDER SEBRING, DAVID KERBOW, SHARON ROLLO, and JOHN Q. EASTON. *Charting Chicago School Reform: Democratic Localism as a Lever for Change*. Dennis Shirley
- 1772 BUHLE, PAUL. *Taking Care of Business: Samuel Gompers, George Meany, Lane Kirkland, and the Tragedy of American Labor*. Kim Scipes
- 1763 BULBECK, CHILLA. *Re-Orienting Western Feminisms: Women's Diversity in a Postcolonial World*. Lourdes Torres
- 300 CANNING, KATHLEEN. *Languages of Labor and Gender: Female Factory Work in Germany, 1850-1914*. Julia Kirk Blackwelder
- 900 CARRIER, JAMES G., and DANIEL MILLER, eds. *Virtualism: A New Political Economy*. Bruce G. Carruthers
- 1207 CARRUTHERS, BRUCE G., and TERENCE C. HALLIDAY. *Rescuing Business: The Making of Corporate Bankruptcy Law in England and the United States*. John Henry Schlegel
- 1228 CARTER, JOHN, ed. *Postmodernity and the Fragmentation of Welfare*. C. G. A. Bryant
- 313 CASPER, MONICA J. *The Making of the Unborn Patient: A Social Anatomy of Fetal Surgery*. Charis Thompson
- 600 CASSELL, JOAN. *The Woman in the Surgeon's Body*. Christine Williams
- 1484 CHAFETZ, JANET SALTZMAN. *Handbook of the Sociology of Gender*. Myra Marx Ferree
- 1810 CHALVON-DEMERSAY, SABINE. *A Thousand Screenplays: The French Imagination in a Time of Crisis*. S. Craig Watkins
- 1797 CHAVES, MARK. *Ordaining Women: Culture and Conflict in Religious Organizations*. R. Stephen Warner
- 1812 CLARKE, LEE. *Mission Improbable: Using Fantasy Documents to Tame Disaster*. Peter K. Manning
- 1828

- 872 CLARKE, ADELE E. *Disciplining Reproduction: Modernity, American Life Sciences, and "the Problems of Sex."* Steven Epstein
- 1242 COHEN, CATHY J. *The Boundaries of Blackness: AIDS and the Breakdown of Black Politics.* Sharon Hicks-Bartlett
- 862 COLLIER, JANE FISHBURNE. *From Duty to Desire: Remaking Families in a Spanish Village.* Arlene Skolnick
- 558 COLLINS, PATRICIA HILL. *Fighting Words: Black Women and the Search for Justice.* Patricia Ticineto Clough
- 528 COLLINS, RANDALL. *The Sociology of Philosophies: A Global Theory of Intellectual Change.* Andrew Abbott
- 1791 CONLEY, DALTON. *Being Black, Living in the Red: Race, Wealth, and Social Policy in America.* David A. Reingold
- 568 COOPER, DAVINA. *Governing Out of Order: Space, Law and the Politics of Belonging.* Neil Brenner
- 1490 CORNELIUS, WAYNE A., TODD A. EISENSTADT, and JANE HINDLEY, eds. *Subnational Politics and Democratization in Mexico.* Jonathan Fox
- 290 CROSS, JOHN C. *Informal Politics: Street Vendors and the State in Mexico City.* Lauren Benton
- 1518 CRUZ, JON. *Culture on the Margins: The Black Spiritual and the Rise of American Cultural Interpretation.* Michael Rogin
- 874 DAVIDSON, JULIA O'CONNELL. *Prostitution, Power and Freedom.* Lynne Segal
- 594 DEBELJAK, ALES. *Reluctant Modernity: The Institution of Art and Its Historical Forms.* Albert J. Bergesen
- 268 DEOLLOS, IONE Y. *On Becoming Homeless: The Shelterization Process for Homeless Families.* Joanna M. Badagliacco
- 552 DODSON, LISA. *Don't Call Us Out of Name: The Untold Lives of Women and Girls in Poor America.* Ruth Horowitz
- 265 DRAKEFORD, MARK. *Social Movements and Their Supporters: The Green Shirts in England.* Ira Silver
- 1222 DUNCAN, CYNTHIA M. *Worlds Apart: Why Poverty Persists in Rural America.* David Brown
- 576 EDLES, LAURA DESFOR. *Symbol and Ritual in the New Spain: The Transition to Democracy after Franco.* Jan Kubik
- 566 ELIASOPH, NINA. *Avoiding Politics: How Americans Produce Apathy in Everyday Life.* Joshua Gamson

- 857 ENARSON, ELAINE, and BETTY HEARN MORROW, eds. *The Gendered Terrain of Disaster: Through Women's Eyes*. Susan Lovegren Bosworth
 - 1206 EPP, CHARLES R. *The Rights Revolution: Lawyers, Activists, and Supreme Courts in Comparative Perspective*. Sidney Tarrow
 - 237 ESPELAND, WENDY NELSON. *The Struggle for Water: Politics, Rationality, and Identity in the American Southwest*. Frank Dobbin
 - 1808 ESSIG, LAURIE. *Queer in Russia: A Story of Self, Sex, and the Other*. Annick Prieur
 - 238 EWICK, PATRICIA, and SUSAN S. SILBEY. *The Common Place of Law: Stories from Everyday Life*. Michael McCann
 - 588 EYERMAN, RON, and ANDREW JAMISON. *Music and Social Movements: Mobilizing Traditions in the Twentieth Century*. Scott A. Hunt
 - 587 FARLEY, JOHN E. *Earthquake Fears, Predictions, and Preparations in Mid-America*. Betty Hearn Morrow
 - 303 FAVELL, ADRIAN. *Philosophies of Integration: Immigration and the Idea of Citizenship in France and Britain*. Carl-Ulrik Schierup
 - 255 FENTON, NATALIE, ALAN BRYMAN, and DAVID DEACON, with PETER BIRMINGHAM. *Mediating Social Science*. Wendy A. Wiedenhoft
 - 284 FERRELL, JEFF, and MARK S. HAMM, eds. *Ethnography at the Edge: Crime, Deviance, and Field Research*. Sudhir Venkatesh
 - 549 FINE, MICHELLE, and LOIS WEIS. *The Unknown City: Lives of Poor and Working-Class Young Adults*. Paul A. Jargowsky
 - 1211 FINN, JANET L. *Tracing the Veins: Of Copper, Culture, and Community from Butte to Chuquicamata*. Barbara Sutton
 - 1196 FRANK, ANDRE GUNDER. *ReOrient: Global Economy in the Asian Age*. Christopher Chase-Dunn
 - 1480 FRASER, MARIAM. *Identity without Selfhood: Simone de Beauvoir and Bisexuality*. Janet Wirth-Cauchon
 - 1475 FRIEDKIN, NOAH E. *A Structural Theory of Social Influence*. Stanley Wasserman
 - 1504 FRIEDMAN, GERALD. *State-Making and Labor-Movements: France and the United States, 1876-1914*. Kenneth H. Tucker, Jr.
 - 1506 FRUNDT, HENRY J. *Trade Conditions and Labor Rights: U.S. Initiatives, Dominican and Central American Responses*. Andrew Schrank
- 1830

- 1224 FÜREDI, FRANK. *The Silent War: Imperialism and the Changing Perception of Race*. John Willinsky
- 1514 FURSTENBERG, FRANK F., JR., THOMAS D. COOK, JACQUELYNNE ECCLES, GLEN H. ELDER, JR., and ARNOLD SAMEROFF. *Managing to Make It: Urban Families and Adolescent Success*. Joyce Ladner
- 560 GABRIEL, JOHN. *Whitewash: Racialized Politics and the Media*. Stephen Ellingson
- 878 GAMSON, JOSHUA. *Freaks Talk Back: Tabloid Talk Shows and Sexual Nonconformity*. Margaret Cerullo
- 864 GARFINKEL, IRWIN, SARA S. McLANAHAN, DANIEL MEYER, and JUDITH SELTZER, eds. *Fathers under Fire: The Revolution in Child Support Enforcement*. Frances Goldscheider
- 1226 GILENS, MARTIN. *Why Americans Hate Welfare: Race, Media, and the Politics of Antipoverty Policy*. Sanford F. Schram
- 853 GILL, ANTHONY. *Rendering unto Caesar: The Catholic Church and the State in Latin America*. Paul Christopher Johnson
- 585 GLAZER, PENINA MIGDAL, and MYRON PERETZ GLAZER. *The Environmental Crusaders: Confronting Disaster and Mobilizing Community*. Sherry Cable
- 849 GOLDFARB, JEFFREY C. *Civility and Subversion: The Intellectual Dynamic in Democratic Society*. Neil McLaughlin
- 1765 GREEN, DECEMBER. *Gender Violence in Africa: African Women's Responses*. Jacqueline Bhabha
- 270 GREGORY, STEVEN. *Black Corona: Race and the Politics of Place in an Urban Community*. Mary Pattillo-McCoy
- 578 GUBRIUM, JABER F., and JAMES A. HOLSTEIN. *The New Language of Qualitative Method*. David Silverman
- 1776 GUTHRIE, DOUGLAS. *Dragon in a Three-Piece Suit: The Emergence of Capitalism in China*. Ching Kwan Lee
- 1198 HAAS, ERNST B. *Nationalism, Liberalism, and Progress*. Vol. 1, *The Rise and Decline of Nationalism*. John A. Armstrong
- 896 HALSETH, GREG. *Cottage Country in Transition: A Social Geography of Change and Contention in the Rural-Recreational Countryside*. James R. Hudson
- 892 HAMERMESH, DANIEL S., and FRANK D. BEAN, eds. *Help or Hindrance? The Economic Implications of Immigration for African Americans*. Lisa Catanzarite

- 860 HAMILTON, SARAH. *The Two-Headed Household: Gender and Rural Development in the Ecuadorean Andes*. Amy Lind
- 1200 HANNIGAN, JOHN. *Fantasy City: Pleasure and Profit in the Post-modern Metropolis*. David Grazian
- 1793 HARRIS, FREDRICK C. *Something Within: Religion in African-American Political Activism*. Mary Pattillo-McCoy
- 247 HASHIMOTO, AKIKO. *The Gift of Generations: Japanese and American Perspectives on Aging and the Social Contract*. Emily M. Agree
- 317 HEIMER, CAROL A., and LISA R. STAFFEN. *For the Sake of the Children: The Social Organization of Responsibility in the Hospital and the Home*. Monica J. Casper
- 1232 HEINEMAN, ELIZABETH D. *What Difference Does a Husband Make? Women and Marital Status in Nazi and Postwar Germany*. Lynne Haney
- 1502 HEINZ, WALTER R., ed. *From Education to Work: Cross-National Perspectives*. Kathryn Borman
- 1769 HELD, DAVID, ANTHONY MCGREW, DAVID GOLDBLATT, and JONATHAN PERRATON. *Global Transformations: Politics, Economics and Culture*. Giovanni Arrighi
- 1477 HENDERSON, KATHRYN. *On Line and On Paper: Visual Representations, Visual Culture, and Computer Graphics in Design Engineering*. John Law
- 310 HERBERT, MELISSA S. *Camouflage Isn't Only for Combat: Gender, Sexuality, and Women in the Military*. Rosemary Pringle
- 564 HERBST, SUSAN. *Reading Public Opinion: How Political Actors View the Democratic Process*. William Hoynes
- 596 HERMANOWICZ, JOSEPH C. *The Stars Are Not Enough: Scientists—Their Passions and Professions*. Theodore P. Gerber
- 1786 HERZOG, HANNA. *Gendering Politics: Women in Israel*. Emily Stoper
- 249 HEWITT, JOHN P. *The Myth of Self-Esteem: Finding Happiness and Solving Problems in America*. Colin Wayne Leach
- 1803 IRVINE, LESLIE. *Codependent Forevermore: The Invention of Self in a Twelve Step Group*. Diane Bjorklund
- 889 JACOBSON, MATTHEW FRYE. *Whiteness of a Different Color: European Immigrants and the Alchemy of Race*. Stephen Steinberg

- 306 JANOSKI, THOMAS. *Citizenship and Civil Society: A Framework of Rights and Obligations in Liberal, Traditional, and Social Democratic Regimes*. J. M. Barbalet
- 1492 JAQUETTE, JANE S., and SHARON L. WOLCHIK, eds. *Women and Democracy: Latin America and Central and Eastern Europe*. Carol Mueller
- 281 JENCKS, CHRISTOPHER, and MEREDITH PHILLIPS, eds. *The Black-White Test Score Gap*. George Farkas
- 315 JENKINS, PHILIP. *Moral Panic: Changing Concepts of the Child Molester in Modern America*. Anthony Oberschall
- 592 KAELEBER, LUTZ. *Schools of Asceticism: Ideology and Organization in Medieval Religious Communities*. Ilana F. Silber
- 1795 KATZENSTEIN, MARY FAINSOD. *Faithful and Fearless: Moving Feminist Protest inside the Church and Military*. Mary Bernstein
- 902 KAY, NEIL M. *Pattern in Corporate Evolution*. Christopher N. Pitelis
- 1204 KEIL, ROGER. *Los Angeles: Globalization, Urbanization and Social Struggles*. Jan Lin
- 1471 KESSLER, SUZANNE J. *Lessons from the Intersexed*. Bernice L. Hausman
- 1788 KLATCH, REBECCA E. *A Generation Divided: The New Left, the New Right, and the 1960s*. David S. Meyer
- 298 KLIGMAN, GAIL. *The Politics of Duplicity: Controlling Reproduction in Ceausescu's Romania*. Joanna Goven
- 1472 KNORR-CETINA, KARIN. *Epistemic Cultures: How the Sciences Make Knowledge*. Trevor Pinch
- 540 LAFREE, GARY. *Losing Legitimacy: Street Crime and the Decline of Social Institutions in America*. James F. Short, Jr.
- 294 LAITIN, DAVID D. *Identity in Formation: The Russian-Speaking Populations in the Near Abroad*. Mark R. Beissinger
- 288 LEDENEVA, ALENA V. *Russia's Economy of Favours: Blat, Networking and Informal Exchange*. Federico Varese
- 1498 LEE, CHING KWAN. *Gender and the South China Miracle: Two Worlds of Factory Women*. Diane L. Wolf
- 847 LEROUX, ROBERT. *Histoire et sociologie en France: De l'histoire-science à la sociologie durkheimienne*. Charles Tilly

- 273 LIEBERMAN, ROBERT C. *Shifting the Color Line: Race and the American Welfare State*. Yvonne Zydan
- 574 LIJPHART, AREND, and CARLOW H. WAISMAN, eds. *Institutional Design in New Democracies: Eastern Europe and Latin America*. Geoffrey Pridham
- 545 LIN, JAN. *Reconstructing Chinatown: Ethnic Enclave, Global Change*. Pyong Gap Min
- 904 LIVINGSTONE, D. W. *The Education-Jobs Gap: Underemployment or Economic Democracy*. William DiFazio
- 544 LOFLAND, LYN H. *The Public Realm: Exploring the City's Quintessential Social Territory*. Mark Gottdiener
- 1217 MADSEN, RICHARD. *China's Catholics: Tragedy and Hope in an Emerging Civil Society*. Daniel H. Bays
- 907 MAIONI, ANTONIA. *Parting at the Crossroads: The Emergence of Health Insurance in the United States and Canada*. Brian Gran
- 250 MARGOLIS, DIANE ROTHBARD. *The Fabric of Self: A Theory of Ethics and Emotions*. Rom Harré
- 572 MARX, ANTHONY W. *Making Race and Nation: A Comparison of South Africa, the United States, and Brazil*. David Theo Goldberg
- 266 MCGOVERN, STEPHEN J. *The Politics of Downtown Development: Dynamic Political Cultures in San Francisco and Washington, D.C.* Dingxin Zhao
- 1807 MELODY, M. E., and LINDA M. PETERSON. *Teaching America about Sex: Marriage Guides and Sex Manuals from the Late Victorians to Dr. Ruth*. Julia A. Ericksen
- 536 MESTROVIC, STJEPAN G. *Anthony Giddens: The Last Modernist*. Charles Camic
- 538 MIZRUCHI, SUSAN L. *The Science of Sacrifice: American Literature and Modern Social Theory*. Alan Sica
- 292 MUKERJI, CHANDRA. *Territorial Ambitions and the Gardens of Versailles*. David A. Bell
- 1478 MUTZ, DIANA C. *Impersonal Influence: How Perceptions of Mass Collectives Affect Political Attitudes*. Lynn M. Sanders
- 1512 NELSON, MARGARET K., and JOAN SMITH. *Working Hard and Making Do: Surviving in Small Town America*. Ann R. Tickamyer
- 1236 NETTLETON, SARAH, and JONATHAN WATSON, eds. *The Body in Everyday Life*. Barry Glassner
- 866 NOCK, STEVEN L. *Marriage in Men's Lives*. Linda J. Waite
- 1834

- 302 O'BRIEN, RUTH. *Worker's Paradox: The Republican Origins of New Deal Labor Policy, 1886-1935*. Dan Clawson
- 1230 O'CONNER, JULIA S., ANN SHOLA ORLOFF, and SHEILA SHAVER. *States, Markets, Families: Gender, Liberalism and Social Policy in Australia, Canada, Great Britain and the United States*. Lisa D. Brush
- 296 OFER, DALIA, and LENORE J. WEITZMAN, eds. *Women in the Holocaust*. Joan B. Wolf
- 1215 OI, JEAN C. *Rural China Takes Off: Institutional Foundations of Economic Reform*. Xueguang Zhou
- 1221 OWENSBY, BRIAN P. *Intimate Ironies: Modernity and the Making of Middle-Class Lives in Brazil*. Miguel Angel Centeno
- 1767 PALUMBO-LIU, DAVID. *Asian/American: Historical Crossings of a Racial Frontier*. Yen Le Espiritu
- 257 PASSAS, NIKOS, and ROBERT AGNEW, eds. *The Future of Anomie Theory*. John H. Gagnon
- 876 PHOENIX, JOANNA. *Making Sense of Prostitution*. Julia O'Connell Davidson
- 1482 PRELL, RIV-ELLEN. *Fighting to Become Americans: Jews, Gender, and the Anxiety of Assimilation*. Deborah A. Gerson
- 1238 PRIEUR, ANNICK. *Mema's House, Mexico City: On Transvestites, Queens, and Machos*. Miguel Díaz-Barriga
- 600 PRINGLE, ROSEMARY. *Sex and Medicine: Gender, Power and Authority in the Medical Profession*. Christine Williams
- 1494 PURUSHOTHAMAN, SANGEETHA. *The Empowerment of Women in India: Grassroots Women's Networks and the State*. Nancy A. Naples and Manisha Desai
- 1496 RAY, RAKA. *Fields of Protest: Women's Movements in India*. Verta Taylor
- 884 RAZ, AVIAD E. *Riding the Black Ship: Japan and Tokyo Disneyland*. Gary Alan Fine
- 311 RISMAN, BARBARA J. *Gender Vertigo: American Families in Transition*. Anita Ilta Garey
- 1779 SASSEN, SASKIA. *Guests and Aliens*. Charles Tilly
- 570 SCHAFER, FREDERIC C. *Democracy in Translation: Understanding Politics in an Unfamiliar Culture*. Jeffrey W. Rubin
- 1782 SHAIN, YOSHI. *Marketing the American Creed Abroad: Diasporas in the U.S. and their Homelands*. Sarah V. Wayland

- 846 SHAPIRO, GILBERT, and JOHN MARKOFF. *Revolutionary Demands: A Content Analysis of the Cahiers de Doleances of 1789*. John A. Hall
- 548 SHARFF, JAGNA WOJCICKA. *King Kong on 4th Street: Families and the Violence of Poverty on the Lower East Side*. Ida Susser
- 1239 SHARP, ELAINE B., ed. *Culture Wars and Local Politics*. Rhys H. Williams
- 279 SHAVIT, YOSSEI, and WALTER MÜLLER, eds., with the editorial assistance of Clare Tame. *From School to Work: A Comparative Study of Educational Qualifications and Occupational Destinations*. Alan C. Kerckhoff
- 245 SHWEDER, RICHARD A., ed. *Welcome to Middle Age! (And Other Cultural Fictions)*. John Modell
- 868 SILVERBERG, HELENE, ed. *Gender and American Social Science: The Formative Years*. Patricia Madoo Lengermann
- 579 SIMPSON, BOB. *Changing Families: An Ethnographic Approach to Divorce and Separation*. Joseph Hopper
- 308 SKEGGS, BEVERLEY. *Formations of Class and Gender: Becoming Respectable*. Carolyn Steedman
- 554 SLEVIN, KATHLEEN F., and C. RAY WINGROVE. *From Stumbling Blocks to Stepping Stones: The Life Experiences of Fifty Professional African American Women*. Eleanor Palo Stoller
- 286 SMITH, MARK M. *Mastered by the Clock: Time, Slavery, and Freedom in the American South*. Tukufu Zuberi
- 241 SMITH, TOBY M. *The Myth of Green Marketing: Tending Our Goats at the Edge of Apocalypse*. John C. Leggett
- 1500 SOLINGER, DOROTHY J. *Contesting Citizenship in Urban China: Peasant Migrants, the State, and the Logic of the Market*. Zai Liang
- 556 ST. JEAN, YANICK, and JOE R. FEAGIN. *Double Burden: Black Women and Everyday Racism*. Lynda F. Dickson
- 1244 STERK, CLAIRE E. *Fast Lives: Women Who Use Crack Cocaine*. Kathryn J. Fox
- 1209 STORPER, MICHAEL, and ROBERT SALAIS. *Worlds of Production: The Action Frameworks of the Economy*. Gerald Berk
- 533 STRENSKI, IVAN. *Durkheim and the Jews of France*. Pierre Birnbaum

- 1486 SZELÉNYI, SZONJA. *Equality by Design: The Grand Experiment in Destratification in Socialist Hungary*. Petr Mateju
- 894 TOLNAY, STEWART E. *The Bottom Rung: African American Family Life on Southern Farms*. Steven Ruggles
- 1469 TREICHLER, PAULA A. *How to Have Theory in an Epidemic: Cultural Chronicles of AIDS*. Lisa Jean Moore
- 1213 TSUTSUI, WILLIAM M. *Manufacturing Ideology: Scientific Management in Twentieth-Century Japan*. Gary Herrigel
- 1219 TUAN, MIA. *Forever Foreigners or Honorary Whites? The Asian Ethnic Experience Today*. Grace Kao
- 855 VANCE, LAURA L. *Seventh-Day Adventism in Crisis: Gender and Sectarian Change in an Emerging Religion*. John P. Bartkowski
- 1508 VARANO, CHARLES S. *Forced Choices: Class, Community, and Worker Ownership*. Vicki Smith
- 890 VICKERMAN, MILTON. *Crosscurrents: West Indian Immigrants and Race*. Vilna Bashi
- 851 VISWANATHAN, GAURI. *Outside the Fold: Conversion, Modernity, and Belief*. Gerald James Larson
- 1774 WARDELL, MARK, THOMAS L. STEIGER, and PETER MEIKSINS, eds. *Rethinking the Labor Process*. Martin Parker
- 263 WARNER, R. STEPHEN, and JUDITH G. WITTNER, eds. *Gatherings in Diaspora: Religious Communities and the New Immigration*. Matthew P. Lawson
- 887 WATKINS, S. CRAIG. *Representing: Hip Hop Culture and the Production of Black Cinema*. Jon Cruz
- 1784 WEDEEN, LISA. *Ambiguities of Domination: Politics, Rhetoric, and Symbols in Contemporary Syria*. George Steinmetz
- 898 WELLMAN, BARRY, ed. *Networks in the Global Village: Life in Contemporary Communities*. Larissa Adler Lomnitz
- 542 WESTERVELT, SAUNDRA DAVIS. *Shifting the Blame: How Victimization Became a Criminal Defense*. Ross Macmillan
- 558 WESTON, KATH. *Long Slow Burn: Sexuality and Social Science*. Patricia Ticineto Clough
- 270 WILSON, WILLIAM H. *Hamilton Park: A Planned Black Community in Dallas*. Mary Pattillo-McCoy
- 1202 WOODIWISS, ANTHONY. *Globalisation, Human Rights, and Labour Law in Pacific Asia*. Pheng Cheah

American Journal of Sociology

- 1801 WUTHNOW, ROBERT. *After Heaven: Spirituality in America since the 1950s*. Peter Dobkin Hall
- 1516 YANG, FENGGANG. *Chinese Christians in America: Conversion, Assimilation, and Adhesive Identities*. Rebecca Kim and Min Zhou
- 1519 ZOLOV, ERIC. *Refried Elvis: The Rise of the Mexican Counterculture*. Pablo Vila

BOOK REVIEWERS

- ANDREW ABBOT, 528
 EMILY M. AGREE, 247
 JOHN A. ARMSTRONG, 1198
 GIOVANNI ARRIGHI, 1769
 JOANNA M. BADAGLIACCO, 268
 WILLIAM SIMS BAINBRIDGE, 261
 J. M. BARBALET, 306
 JOHN P. BARTKOWSKI, 855
 VILNA BASHI, 890
 DANIEL H. BAYS, 1217
 MARK R. BEISSINGER, 294
 DAVID A. BELL, 292
 LAUREN BENTON, 290
 ALBERT J. BERGESEN, 594
 GERALD BERK, 1209
 MARY BERNSTEIN, 1795
 JACQUELINE BHABHA, 1765
 PIERRE BIRNBAUM, 533
 DIANE BJORKLUND, 1803
 JULIA KIRK BLACKWELDER, 300
 KATHRYN BORMAN, 1502
 SUSAN LOVEGREN BOSWORTH, 857
 KEVIN BREAULT, 243
 NEIL BRENNER, 568
 DAVID BROWN, 1222
 LISA D. BRUSH, 1230
 C. G. A. BRYANT, 1228
 PETER BURKE, 883
 SHERRY CABLE, 585
 CHARLES CAMIC, 536
 BRUCE G. CARRUTHERS, 900
 MONICA J. CASPER, 317
 LISA CATANZARITE, 892
 MIGUEL ANGEL CENTENO, 1221
 MARGARET CERULLO, 878
 CHRISTOPHER CHASE-DUNN, 1196
 PHENG CHEAH, 1202
 NICHOLAS A. CHRISTAKIS, 1246
 DAN CLAWSON, 302
 PATRICIA TICINETO CLOUGH, 558
 DALTON CONLEY, 1790
 MAXINE CRAIG, 1805
 JON CRUZ, 887
 JULIA O'CONNELL DAVIDSON, 876
 MANISHA DESAI, 1494
 MIGUEL DÍAZ-BARRIGA, 1238
 LYNDA F. DICKSON, 556
 WILLIAM DiFAZIO, 904
 FRANK DOBBIN, 237
 STEPHEN ELLINGSON, 560
 STEVEN EPSTEIN, 872
 JULIA A. ERICKSEN, 1807
 YEN LE ESPIRITU, 1767
 GEORGE FARKAS, 281
 MYRA MARX FERREE, 1484
 GARY ALAN FINE, 884
 JONATHAN FOX, 1490
 KATHRYN J. FOX, 1244
 RUTH FRANKENBERG, 562
 STEPHEN FUCHS, 1814
 JOHN H. GAGNON, 257, 870
 JOSHUA GAMSON, 566
 ANITA ILTA GAREY, 311
 THEODORE P. GERBER, 596
 DI'BORAH A. GERSON, 1482
 ANDREAS GLAESER, 1488
 BARRY GLASSNER, 1236
 DAVID THEO GOLDBERG, 572
 FRANCES GOLDSCHIEDER, 864
 JACK A. GOLDSTONE, 531
 MARK GOTTDIENER, 544
 MICHAEL R. GOTTFREDSON, 283
 JOANNA GOVEN, 298
 BRIAN GRAN, 907
 DAVID GRAZIAN, 1200
 JOHN A. HALL, 846
 PETER DOBKIN HALL, 1801
 MAUREEN T. HALLINAN, 583
 LYNNE HANEY, 1232
 ROM HARRÉ, 250
 BERNICE L. HAUSMAN, 1471
 GARY HERRIGEL, 1213
 SHARON HICKS-BARTLETT, 1242
 E. J. HOBBSBAM, 844
 RANDY HODSON, 1770
 JOSEPH HOPPER, 579
 RUTH HOROWITZ, 552
 WILLIAM HOYNES, 564
 JAMES R. HUDSON, 896
 SCOTT A. HUNT, 588
 THOMAS JANOSKI, 909
 PAUL A. JARGOWSKY, 549
 PAUL CHRISTOPHER JOHNSON, 853
 GRACE KAO, 1219
 ALAN C. KERCKHOFF, 279
 REBECCA KIM, 1516
 JOHN A. KOVACH, 277
 JAN KUBIK, 576
 JOYCE LADNER, 1514
 GERALD JAMES LARSON, 851

American Journal of Sociology

- JOHN LAW, 1477
MATTHEW P. LAWSON, 263
COLIN WAYNE LEACH, 249
CHING KWAN LEE, 1776
JOHN C. LEGGETT, 241
PATRICIA MADOO LENG-ERMANN, 868
ZAI LIANG, 1500
JAN LIN, 1204
AMY LIND, 860
LARISSA ADLER LOMNITZ, 898
ROSS MACMILLAN, 542
DAVID R. MAINES, 880
PETER K. MANNING, 1812
PETR MATEJU, 1486
DOUGLAS W. MAYNARD, 598
MICHAEL MCCANN, 238
NEIL MCLAUGHLIN, 849
DAVID S. MEYER, 1788
PYONG GAP MIN, 545
JOHN MODELL, 245
LISA JEAN MOORE, 1469
BETTY HEARN MORROW, 587
MARY PATTILLO-MCCOY, 270, 1793
CAROL MUELLER, 1492
NANCY A. NAPLES, 1494
ANTHONY OBERSCHALL, 315
MARTIN PARKER, 1774
MARY PATTILLO-MCCOY, 270, 1793
TREVOR PINCH, 1472
CHRISTOPHER N. PITELIS, 902
GEOFFREY PRIDHAM, 574
ANNICK PRIEUR, 1808
ROSEMARY PRINGLE, 310
DAVID A. REINGOLD, 1791
MICHAEL ROGIN, 1518
BARBARA KATZ ROTHMAN, 1234
JEFFREY W. RUBIN, 570
STEVEN RUGGLES, 894
LYNN M. SANDERS, 1478
CARL-ULRIK SCHIERUP, 303
JOHN HENRY SCHLEGEL, 1207
KIM SCIPES, 1772
SANFORD F. SCHRAM, 1226
ANDREW SCHRANK, 1506
LYNNE SEGAL, 874
DENNIS SHIRLEY, 275
JAMES F. SHORT, JR., 540
ALAN SICA, 538
ILANA F. SILBER, 592
SUSAN SILBEY, 252
IRA SILVER, 265
DAVID SILVERMAN, 578
ARLENE SKÖLNICK, 862
VICKI SMITH, 1508
CAROLYN STEEDMAN, 308
STEPHEN STEINBERG, 889
GEORGE STEINMETZ, 1784
MITCHELL L. STEVENS, 1799
ARTHUR L. STINCHCOMBE, 1761
ELEANOR PALO STOLLER, 554
EMILY STOPER, 1786
IDA SUSSEY, 548
BARBARA SUTTON, 1211
SYDNEY TARROW, 1206
PETER J. TAYLOR, 1780
VERTA TAYLOR, 1496
CHARIS THOMPSON, 313
ANN R. TICKAMYER, 1512
CHARLES TILLY, 847, 1779
EDWARD A. TIRYAKIAN, 258
LOURDES TORRES, 1763
KENNETH H. TUCKER, JR., 1504
SUDHIR VENKATESH, 284
CHARLES S. VARANO, 1510
FEDERICO VARESE, 288
PABLO VILA, 1519
LINDA J. WAITE, 581, 866
SYLVIA WALBY, 859
R. STEPHEN WARNER, 1797
STANLEY WASSERMAN, 1475
S. CRAIG WATKINS, 1810
SARAH V. WAYLAND, 1782
WENDY A. WIEDENHOFT, 255
CHRISTINE WILLIAMS, 600
RHYS H. WILLIAMS, 1239
JOHN WILLINSKY, 1224
JANET WIRTH-CAUCHON, 1480
DIANE L. WOLF, 1498
JOAN B. WOLF, 296
DINGXIN ZHAO, 266
MIN ZHOU, 1516
XUEGUANG ZHOU, 1215
SANDRA L. ZIMDARS-SWARTZ, 590
TUKUFU ZUBERI, 286
YVONNE ZYLAN, 273